

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 10:10:37 ; Search time 35.2258 Seconds
(without alignments)
330.836 Million cell updates/sec

Title: US-10-071-411A-6

Perfect score: 21

Sequence: 1 tcatgtatcccaattagagact 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*

4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/prodata/2/ina/PCUS_COMB.seq:*

6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	2189	3	US-08-846-020A-1
2	19.4	92.4	2189	4	US-09-617-871-1
3	16.2	77.1	1431	4	US-09-328-352-901
4	16.2	77.1	72604	4	US-09-268-992-7
5	16.2	77.1	72604	4	US-09-657-474-7
6	15.8	75.2	168575	4	US-09-426-290-1
7	15.4	73.3	5864	3	US-08-894-440-4
8	15.4	73.3	5864	4	US-09-458-093-4
9	15.4	73.3	5865	4	US-09-430-497A-1
10	15.4	73.3	14113	3	US-09-223-134-1
11	15.4	73.3	14113	3	US-08-992-801-1
12	15.4	73.3	14113	3	US-09-223-535-1
13	15.4	73.3	14194	4	US-09-577-424-3
14	15.4	73.3	15397	2	US-08-673-768-1
15	15.4	73.3	15397	2	US-08-673-768-1
16	15.2	72.4	1812	4	US-09-543-681A-2345
17	15.2	72.4	2070	4	US-09-620-312D-958
18	15.2	72.4	2734	4	US-08-375-134-13
19	15.2	72.4	2734	5	PCT-US95-15263-13
20	14.8	70.5	257	3	US-08-956-171E-3108
21	14.8	70.5	536	3	US-09-147-928-3
22	14.8	70.5	581	3	US-08-998-416-743
23	14.8	70.5	787	4	US-09-976-594-763
24	14.8	70.5	867	4	US-09-543-681A-1203
25	14.8	70.5	1242	3	US-09-147-928-1
26	14.8	70.5	1320	4	US-08-956-171E-80
27	14.8	70.5	1323	1	US-08-307-499-36

c	28	14.8	70.5	1323	3	US-09-299-268-36	Sequence 36, Appl
	29	14.8	70.5	1351	1	US-07-816-283-5	Sequence 5, Appl
	30	14.8	70.5	1351	1	US-08-417-103-5	Sequence 5, Appl
	31	14.8	70.5	1351	4	US-09-016-434-1303	Sequence 1303, Ap
c	32	14.8	70.5	1372	4	US-09-530-836-1	Sequence 1, Appl
	33	14.8	70.5	1849	2	US-08-912-227-1	Sequence 1, Appl
	34	14.8	70.5	1849	4	US-09-513-584-1	Sequence 1, Appl
	35	14.8	70.5	1849	4	US-09-345-790-1	Sequence 1, Appl
c	36	14.8	70.5	2423	3	US-08-714-918-86	Sequence 86, Appl
	37	14.8	70.5	2423	3	US-09-265-315-86	Sequence 86, Appl
c	38	14.8	70.5	2423	3	US-09-266-417-86	Sequence 86, Appl
	39	14.8	70.5	2423	3	US-09-266-417-86	Sequence 86, Appl
c	40	14.8	70.5	2423	4	US-09-528-705-86	Sequence 86, Appl
	41	14.8	70.5	2423	4	US-09-527-745-86	Sequence 86, Appl
c	42	14.8	70.5	10320	4	US-08-956-171E-143	Sequence 143, App
	43	14.8	70.5	11444	2	US-08-222-617A-26	Sequence 26, Appl
	44	14.8	70.5	12364	2	US-08-222-617A-1	Sequence 1, Appl
	45	14.8	70.5	14176	1	US-08-307-499-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-08-846-020A-1

; Sequence 1, Application US/08846020A

; Patent No. 6050547

; GENERAL INFORMATION:

; APPLICANT: Drazen M.D., Jeffrey M.

; APPLICANT: In M.D., Kwang-Ho

; APPLICANT: Asano M.D., Koichiro

; APPLICANT: Beier, David

; APPLICANT: Grobholz, James

; TITLE OF INVENTION: 5-lipoxygenase Gene Sequence

; TITLE OF INVENTION: Polymorphisms and Their Use in Classifying Patients

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CHOATE, HALL & STEWART

; STREET: 53 State Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109-2891

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/846,020A

; FILING DATE:

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Jarrell Ph.D., Brenda H.

; REGISTRATION NUMBER: 39,223

; REFERENCE/DOCKET NUMBER: 0092662-0012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 248-5000

; TELEFAX: (617) 248 4000

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2189 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: both

; TOPOLOGY: not relevant

; MOLECULE TYPE: DNA (genomic)

; IMMEDIATE SOURCE:

; CLONE: human 5-lipoxygenase gene (GenBank M 38191)

; US-08-846-020A-1

Query Match 92.4%; Score 19.4; DB 3; Length 2189;
Best Local Similarity 95.2%; Pred. No. 0.7;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY      1 TCATGTCATCCCAATTAGAGACT 21
Db      549 TCATGTCATCCGATTAGAGACT 569

RESULT 2
US-09-617-871-1
; Sequence 1, Application US/09617871
; Patent No. 6355434
; GENERAL INFORMATION:
; APPLICANT: Drazén M.D., Jeffrey M.
; APPLICANT: In M.D., Kwang-Ho
; APPLICANT: Asano M.D., Koichiro
; APPLICANT: Beier, David
; APPLICANT: Grobholz, James
; TITLE OF INVENTION: 5-Lipoxygenase Gene Sequence
; TITLE OF INVENTION: Polymorphisms and Their Use in Classifying Patients
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHOATE, HALL & STEWART
; STREET: 53 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2891
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/617,871
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/846,020
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jarrell Ph.D., Brenda H.
; REGISTRATION NUMBER: 39,223
; REFERENCE/DOCKET NUMBER: 0092662-0012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-5000
; TELEFAX: (617) 248 4000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2189 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: human 5-lipoxygenase gene (GenBank M 38191)
US-09-617-871-1

Query Match      92.4%; Score 19.4; DB 4; Length 2189;
Best Local Similarity 95.2%; Pred. No. 0.7;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCATGTCATCCCAATTAGACT 21
Db      549 TCATGTCATCCGATTAGACT 569

RESULT 3
US-09-328-352-901
; Sequence 901, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
```

```
FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 901
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-901

Query Match      77.1%; Score 16.2; DB 4; Length 1431;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TCATGTCATCCCAATTAGACT 21
Db      1235 TCATGTCATCCATTAGCAACT 1255

RESULT 4
US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

Query Match      77.1%; Score 16.2; DB 4; Length 72604;
Best Local Similarity 85.7%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TCATGTCATCCCAATTAGACT 21
Db      20801 TCCTGTATCCCAATAACAGACT 20821

RESULT 5
US-09-657-474-7
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/657,474
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/268,992
```

;/ PRIOR FILING DATE: 1999-03-16
;/ PRIOR APPLICATION NUMBER: 09/236,134
;/ PRIOR FILING DATE: 1999-01-22
;/ PRIOR APPLICATION NUMBER: 60/106,056
;/ PRIOR FILING DATE: 1998-10-28
;/ PRIOR APPLICATION NUMBER: 60/088,312
;/ PRIOR FILING DATE: 1998-06-05
;/ PRIOR APPLICATION NUMBER: 60/078,044
;/ PRIOR FILING DATE: 1998-03-16
;/ NUMBER OF SEQ ID NOS: 84
;/ SOFTWARE: FastSeq for Windows Version 3.0
;/ SEQ ID NO 7
;/ LENGTH: 72604
;/ TYPE: DNA
;/ ORGANISM: Homo sapiens
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: all n positions
;/ OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

Query Match 77.1%; Score 16.2; DB 4; Length 72604;
Best Local Similarity 85.7%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
|||
Db 20801 TCCTGTATCCCAATTAGAGACT 20821

RESULT 6
US-09-426-290-1
;/ Sequence 1, Application US/09426290
;/ Patent No. 6410712
;/ GENERAL INFORMATION:
;/ APPLICANT: Berglind Ran Olafsdottir
;/ APPLICANT: Jeffrey Gulcher
;/ TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
;/ FILE REFERENCE: 2345.2001-000
;/ CURRENT APPLICATION NUMBER: US/09/426,290
;/ CURRENT FILING DATE: 1999-10-25
;/ NUMBER OF SEQ ID NOS: 24
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO 1
;/ LENGTH: 168575
;/ TYPE: DNA
;/ ORGANISM: Homo Sapiens
;/ FEATURE:
;/ NAME/KEY: CDS
;/ LOCATION: (21181)...(21403)
;/ NAME/KEY: CDS
;/ LOCATION: (95252)...(95430)
;/ NAME/KEY: CDS
;/ LOCATION: (101753)...(101996)
;/ NAME/KEY: CDS
;/ LOCATION: (110324)...(110439)
;/ NAME/KEY: CDS
;/ LOCATION: (124058)...(124278)
;/ NAME/KEY: CDS
;/ LOCATION: (127009)...(127130)
;/ NAME/KEY: CDS
;/ LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match 75.2%; Score 15.8; DB 4; Length 168575;
Best Local Similarity 89.5%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGTATCCCAATTAGAGACT 21
|||
Db 86270 ATTTGTCCCAATTAGAGACT 86288

RESULT 7
US-08-894-440-4
;/ Sequence 4, Application US/08894440
;/ Patent No. 6025546
;/ GENERAL INFORMATION:
;/ APPLICANT: PLANT GENETIC SYSTEMS N.V.
;/ TITLE OF INVENTION: Method to obtain male sterile plants
;/ FILE REFERENCE: NM508
;/ CURRENT APPLICATION NUMBER: US/08/894,440
;/ CURRENT FILING DATE: 1997-11-12
;/ NUMBER OF SEQ ID NOS: 4
;/ SOFTWARE: PatentIn Ver. 2.0
;/ SEQ ID NO 4
;/ LENGTH: 5864
;/ TYPE: DNA
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
;/ OTHER INFORMATION: plasmid pTC0113
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((1)..(25))
;/ OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((98)..(330))
;/ OTHER INFORMATION: region containing polyadenylation signal of gene 7
;/ OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((331)..(882))
;/ OTHER INFORMATION: region coding for phosphinothricin acetyl
;/ OTHER INFORMATION: transferase (bar)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((883)..(2608))
;/ OTHER INFORMATION: promoter of small subunit gene of Rubisco of
;/ OTHER INFORMATION: Arabidopsis (Pssu)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((2659)..(3031))
;/ OTHER INFORMATION: region containing polyadenylation signal of
;/ OTHER INFORMATION: nopaline synthase Gene of Agrobacterium T-DNA
;/ OTHER INFORMATION: (3'nos)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((3032)..(3367))
;/ OTHER INFORMATION: region coding for barnase of Bacillus
;/ OTHER INFORMATION: amyloliquefaciens
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((3368)..(4877))
;/ OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
;/ OTHER INFORMATION: tabacum (PTA29)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (4924)..(5216)
;/ OTHER INFORMATION: promoter of nopaline synthase gene of
;/ OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (5217)..(5489)
;/ OTHER INFORMATION: region coding for barstar of Bacillus
;/ OTHER INFORMATION: amyloliquefaciens
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (5490)..(5765)
;/ OTHER INFORMATION: region containing polyadenylation signal of gene 7
;/ OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: Complement((5840)..(5864))
;/ OTHER INFORMATION: left border of Agrobacterium T-DNA

```
US-08-894-440-4
Query Match          73.3%; Score 15.4; DB 3; Length 5864;
Best Local Similarity 94.1%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTATCCAATTAGAGACT 21
|||||
Db 5177 GTATCCAATTAGAGTCT 5193

RESULT 8
US-09-458-093-4
; Sequence 4, Application US/09458093
; Patent No. 6344602
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/09/458,093
; CURRENT FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 08/894,440
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTCO113
; NAME/KEY: misc feature
; LOCATION: Complement(1)..(25)
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; NAME/KEY: misc feature
; LOCATION: Complement(98)..(330)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; NAME/KEY: misc feature
; LOCATION: Complement(331)..(882)
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; NAME/KEY: misc feature
; LOCATION: Complement(883)..(2608)
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; NAME/KEY: misc feature
; LOCATION: Complement(2659)..(3031)
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; NAME/KEY: misc feature
; LOCATION: Complement(3032)..(3367)
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; NAME/KEY: misc feature
; LOCATION: Complement(3368)..(4877)
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; NAME/KEY: misc feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; NAME/KEY: misc feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; NAME/KEY: misc feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; NAME/KEY: misc feature

; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-09-458-093-4
Query Match          73.3%; Score 15.4; DB 4; Length 5864;
Best Local Similarity 94.1%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTATCCAATTAGAGACT 21
|||||
Db 5177 GTATCCAATTAGAGTCT 5193

RESULT 9
US-09-430-497A-1
; Sequence 1, Application US/09430497A
; Patent No. 6509516
; GENERAL INFORMATION:
; APPLICANT: WESTON, Brigitte
; APPLICANT: DE BEUCKELEER, Marc
; TITLE OF INVENTION: MALE-STERILE BRASSICA PLANTS AND METHOD FOR PRODUCING
; TITLE OF INVENTION: SAME
; FILE REFERENCE: 51412-2020
; CURRENT APPLICATION NUMBER: US/09/430,497A
; CURRENT FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pCO113
US-09-430-497A-1
Query Match          73.3%; Score 15.4; DB 4; Length 5865;
Best Local Similarity 94.1%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTATCCAATTAGAGACT 21
|||||
Db 5178 GTATCCAATTAGAGTCT 5194

RESULT 10
US-09-223-134-1/c
; Sequence 1, Application US/09223134
; Patent No. 6057490
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Uknes, Scott
; APPLICANT: Ward, Eric
; APPLICANT: Delaney, Terry
; APPLICANT: Lawton, Ray
; APPLICANT: Weymann, Kris
; APPLICANT: Steiner, Henry-York
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; TITLE OF INVENTION: Into Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6057490artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```


;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/223,134
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/992,801
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; APPLICATION NUMBER: US/08/165,248
;; FILING DATE: 10-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/002,285
;; FILING DATE: 08-JAN-1993
;; TELEPHONE: 919-541-8587
;; NAME: Meigs, J. Timothy
;; REGISTRATION NUMBER: 38,241
;; REFERENCE/DOCKET NUMBER: CGC1673/CIP3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919-541-8587
;; TELEFAX: 919-541-8689
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1413 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: other nucleic acid
;; DESCRIPTION: /desc = "PR-1/luc construct"
US-09-223-134-1

Query Match 73.3%; Score 15.4; DB 3; Length 14113;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTATCCCAATTAGACT 21
DB 57 GTATCCCAATTAGACT 41

RESULT 11
US-09-223-134-1/c
; Sequence 1, Application US/08992801
; Patent No. 6107544
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Uknes, Scott
; APPLICANT: Ward, Eric
; APPLICANT: Delaney, Terry
; APPLICANT: Lawton, Kay
; APPLICANT: Weymann, Kris
; APPLICANT: Steiner, Henry-York
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; TITLE OF INVENTION: Into Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6107544artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,801
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/648,949

;; FILING DATE: 16-MAY-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/165,248
;; FILING DATE: 10-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/002,285
;; FILING DATE: 08-JAN-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meigs, J. Timothy
;; REGISTRATION NUMBER: 38,241
;; REFERENCE/DOCKET NUMBER: CGC1673/CIP3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919-541-8587
;; TELEFAX: 919-541-8689
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1413 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: other nucleic acid
;; DESCRIPTION: /desc = "PR-1/luc construct"
US-08-992-801-1

Query Match 73.3%; Score 15.4; DB 3; Length 14113;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTATCCCAATTAGACT 21
DB 57 GTATCCCAATTAGACT 41

RESULT 12
US-09-223-535-1/c
; Sequence 1, Application US/09223535
; Patent No. 6232525
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Uknes, Scott
; APPLICANT: Ward, Eric
; APPLICANT: Delaney, Terry
; APPLICANT: Lawton, Kay
; APPLICANT: Weymann, Kris
; APPLICANT: Steiner, Henry-York
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; TITLE OF INVENTION: Into Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6232525artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/223,535
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/992,801
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,248
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/002,285

2

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15397 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-673-768-1

```

```

Query Match      73.3%; Score 15.4; DB 2; Length 15397;
Best Local Similarity 94.1%; Pred.No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5 GTATCCAATTAGAGACT 21
Db      4575 GTATCCAATTAGAGTCT 4559

```

```

Search completed: May 7, 2004, 11:56:24
Job time : 37.2258 secs

```

Mon May 10 11:38:45 2004

us-10-071-411a-6.rst

Page 8

Db 286 ATGTATCCCAATTAGAGA 270

Search completed: May 7, 2004, 11:54:29
Job time : 1442.5 secs

All repeats were identified using RepeatMasker:
Smit A E & Green P (1996-1997)

REFERENCE 3 (bases 1 to 65598)

Only the middle 65.6 kilobases of this clone are being submitted. The remainder is overlapped either by accession number AC068014 [WIGR project L10137] or accession number AC087222 [WIGR project L12027].

```
1. .65598
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
```

```
/clone lib="RPCT-11 Human Male BAC"
```

repeat_region	complement(264, .677)
repeat_region	/rpt family="MLT2D"
repeat_region	complement(268, .2201)
repeat_region	/rpt family="MLT1F2"
repeat_region	complement(2206, .2484)
repeat_region	/rpt family="MLT1F2"
repeat_region	complement(2645, .2723)
repeat_region	/rpt family="MIR"
repeat_region	complement(2576, .3121)
repeat_region	/rpt family="MIR"
repeat_region	8012, .8252
repeat_region	/rpt family="AluJo"
repeat_region	complement(10348, .10641)
repeat_region	/rpt family="AluSc"
repeat_region	complement(10643, .10923)
repeat_region	/rpt family="AluJo"
repeat_region	11073, .11097
repeat_region	/rpt family="(TC)n"
repeat_region	complement(12317, .12481)
repeat_region	/rpt family="AluSg/x"
repeat_region	complement(12987, .13147)
repeat_region	/rpt family="MER106A"
repeat_region	complement(13703, .14336)
repeat_region	/rpt family="LMB7"
repeat_region	complement(14337, .14633)
repeat_region	/rpt family="AluSg"
repeat_region	complement(14634, .14947)
repeat_region	/rpt family="LMB7"
repeat_region	complement(14953, .15099)
repeat_region	/rpt family="AluJo"
repeat_region	complement(15100, .15403)
repeat_region	/rpt family="AluJb"
repeat_region	complement(15404, .15533)

JOURNAL
Submitted (13-JAN-2003) Whitehead Institute/MIT Center
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

repeat_region
/rpt_family="MER106A"
complement(13703..14336)
/rpt_family="L1MB7"
repeat_region
complement(14337..14633)
/rpt_family="AluSq"
repeat_region
complement(14634..14947)
/rpt_family="L1MB7"
repeat_region
complement(14953..15099)
/rpt_family="AluJo"
repeat_region
complement(15100..15403)
/rpt_family="AluJb"
repeat_region
complement(15404..15533)

```

```

repeat_region      /rpt family="AluJo"
                   complement(15543..15706)
repeat_region      /rpt family="L1MB7"
                   16293..16418
repeat_region      /rpt family="FLAM A"
                   16592..16894
repeat_region      /rpt family="AluSx"
                   17292..17458
repeat_region      /rpt family="AluJo"
                   complement(17782..18084)
repeat_region      /rpt family="AluSx"
                   complement(18275..18480)
repeat_region      /rpt family="L1M4c"
                   complement(18490..18746)
repeat_region      /rpt family="L1M4c"
                   complement(18747..19045)
repeat_region      /rpt family="AluDb"
                   complement(19046..19185)
repeat_region      /rpt family="L1M4c"
                   19693..19785
repeat_region      /rpt family="AT rich"
                   complement(20643..20801)
repeat_region      /rpt family="L2"
                   complement(21504..21590)
repeat_region      /rpt family="L1MB8"
                   complement(21607..21997)
repeat_region      /rpt family="L1MB4"
                   complement(21998..22312)
repeat_region      /rpt family="AluY"
                   complement(22319..22486)
repeat_region      /rpt family="FAM"
                   complement(22487..22676)
repeat_region      /rpt family="L1MB4"
                   complement(23455..23641)

```

Query Match 84.8%; Score 17.8; DB 9; Length 65598;
 Best Local Similarity 90.5%; Pred. NO. 1.2e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTCATCCCAATAGAGACT 21

Db 14896 TCATATATCCCAATAAGAGACT 14876

RESULT 10
 AL139393/c
 LOCUS Human DNA sequence from clone RP3-428L16 on chromosome 6q26-27, linear PRI 30-NOV-2000
 DEFINITION complete sequence.
 ACCESSION AL139393
 VERSION AL139393.13 GI:11414475
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 85624)
 AUTHORS Blakey, S.
 TITLE Direct Submission
 JOURNAL Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerquest@sanger.ac.uk
 On Nov 28, 2000 this sequence version replaced gi:1128544.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated repeat sequence elements. Where the sequence is

ambiguous, there is an annotation using the 'unsure' feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr6
 RP3-428L16 is from the library RPCI-3 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: pCYPAC2

This sequence is the entire insert of clone RP3-428L16 The true
 right end of clone RP11-235G24 is at 64894 in this sequence.

FEATURES	Location/Qualifiers
source	1..85624
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="6"
	/map="q26-27"
	/clone="RP3-428L16"
	/clone_lib="RPCI-3"
repeat_region	1..57
	/note="Alu repeat: matches 231..291 of consensus"
repeat_region	1248..1439
	/note="HAL1 repeat: matches 1406..1592 of consensus"
repeat_region	1588..1645
	/note="6 copies 13 mer 74% conserved"
repeat_region	1742..1916
	/note="L1M3 repeat: matches 5435..5608 of consensus"
repeat_region	1917..2227
	/note="AluJo repeat: matches 1..311 of consensus"
repeat_region	2228..7263
	/note="L1M3 repeat: matches 678..5435 of consensus"
repeat_region	7245..7997
	/note="L1M3a repeat: matches 202..982 of consensus"
repeat_region	7997..8211
	/note="L1M4 repeat: matches 4847..5134 of consensus"
repeat_region	8219..8472
	/note="HAL1 repeat: matches 548..834 of consensus"
repeat_region	8667..8885
	/note="L1MC4 repeat: matches 7533..7757 of consensus"
repeat_region	8894..8951
	/note="29 copies 2 mer ac 82% conserved"
repeat_region	9060..9354
	/note="AluSc repeat: matches 8..303 of consensus"
repeat_region	9363..9497
	/note="FLAM A repeat: matches 1..131 of consensus"
repeat_region	10317..10356
	/note="20 copies 2 mer ca 87% conserved"
repeat_region	10762..11068
	/note="AluDb repeat: matches 1..306 of consensus"
repeat_region	11204..11667
	/note="L2 repeat: matches 2244..2710 of consensus"
misc_feature	complement(11526..12078)
	/note="match: GSS: Em:AQ552369"
repeat_region	11715..11991
	/note="L2 repeat: matches 1809..2104 of consensus"
misc_feature	12081..12662
	/note="match: GSS: Em:AQ389233"
repeat_region	13932..14258
	/note="AluJo repeat: matches 1..311 of consensus"
repeat_region	14438..14736
	/note="AluJo repeat: matches 1..301 of consensus"
repeat_region	14873..15174
	/note="AluSx repeat: matches 11..312 of consensus"
repeat_region	15565..15876
	/note="AluDb repeat: matches 8..312 of consensus"
repeat_region	16789..16812
	/note="12 copies 2 mer tt 100% conserved"

```
misc_feature
complement(17714..18094)
/note="match: GSS: Em:AQ038730"
repeat_region
18567..19061
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
19163..19198
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
19266..19411
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
19534..19605
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
19759..20025
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
20240..20274
/note="L2 repeat: matches 1894..2417 of consensus"
misc_feature
complement(20571..20954)
/note="match: STS: Em:G17060"
repeat_region
20823..21048
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
21064..21607
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
21620..21979
/note="L2 repeat: matches 1894..2417 of consensus"
misc_feature
21639..22101
/note="match: GSS: Em:AQ410503"
repeat_region
22107..22366
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
23046..24355
/note="L2 repeat: matches 1894..2417 of consensus"
misc_feature
24450..24784
/note="match: GSS: Em:AQ012538"
repeat_region
25261..25310
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
26330..26427
/note="L2 repeat: matches 1894..2417 of consensus"
misc_feature
27692..28097
/note="match: GSS: Em:AQ015630"
complement(27698..28198)
/note="match: GSS: Em:AQ354193"
28297..28802
/note="match: STS: Em:G58353"
match: GSS: Em:AQ347822"
29477..30101
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
30741..31195
/note="L2 repeat: matches 1894..2417 of consensus"
misc_feature
complement(31484..31806)
/note="match: GSS: Em:AQ044222"
32226..32412
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
33268..33426
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
33456..33489
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
33491..33627
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
34449..34542
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
34643..34790
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
34832..34897
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
34895..34963
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
35207..35487
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
35870..35947
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
35962..36102
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
36103..36395
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
36396..36610
/note="L2 repeat: matches 1894..2417 of consensus"
```

```
repeat_region
37106..37420
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
37687..38041
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
38977..39263
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
39663..39939
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
39941..39984
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
42187..42561
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
43081..43206
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
43217..43768
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
43802..43928
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
43929..44224
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
44225..44483
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
44484..44885
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
44886..45213
/note="L2 repeat: matches 1894..2417 of consensus"
repeat_region
45214..45310
/note="L2 repeat: matches 1894..2417 of consensus"
```

Query Match 84.8%; Score 17.8; DB 9; Length 85624;
Best Local Similarity 90.5%; Pred. NO. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTATCCATTAGAGACT 21

Db 83430 TCATGTATCTAATAAGAGACT 83410

```
RESULT 11
AC005926/c
LOCUS
DEFINITION
AC005926
AC005926.1 GI:4204698
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 95391)
Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z.,
Di.W., Ding,X., Dugan,S., Durbin,J., Forcum,J., Garcia,C.,
Gorell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L.,
Kondeljewski,N., Leal,B., Lichtarge,O., Liu,W., Logan,O., Lu,J.,
Martinez,C., Oswald,C., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,O.,
Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.A.
Direct Submission
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 95391)
Worley,K.C.
Direct Submission
JOURNAL
Submitted (04-NOV-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE
3 (bases 1 to 95391)
Worley,K.C.
Direct Submission
JOURNAL
Submitted (31-JAN-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE
4 (bases 1 to 95391)
Worley,K.C.
```


TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (02-FEB-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 95391)
Worley.K.C.
Direct Submission
Submitted (29-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 95391)
Worley.K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 31, 1999 this sequence version replaced gi:4176310.
INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.

Version: 1.01 xfo.
Location/Qualifiers
1. .95391
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosomes="X"
/clone="GSHB-31214"
1. .111
/note="Overlaps bases 124241 to 124351 in AC005185."
/function="Overlap with adjacent clone AC005185."
complement(509. .547)
/rpt_family="(CA)n"
720. .810
/rpt_family="MER5A"
2055. .2389
/standard_name="AHC.PCR2"
/db_xref="GDB:581666"
complement(2127. .3529)
/gene="DAX-1"
/note="DAX-1=DSS-AHC critical region on X chromosome, gene
1 [human, adrenal hypoplasia congenita, testis, fetal
adrenal gland, mRNA Mutant, 2022 nt], S74720."

FEATURES
Source
misc_feature
repeat_region
repeat_region
STS
gene

2341. .3329
/standard_name="sWXD2736"
/db_xref="GDB:555626"
2481. .2640
/standard_name="2874/2875"
/db_xref="GDB:555638"
2626. .3014
/standard_name="1AaF/1AaR"
/db_xref="GDB:555633"
3801. .3869
/rpt_family="MER5A"
complement(3966. .4043)
/rpt_family="L2"
complement(4119. .4512)
/rpt_family="L1PA15"
complement(4689. .4810)
/rpt_family="(GGAA)n"
4956. .5047
/rpt_family="MIR"
complement(5048. .5488)
/rpt_family="MLT1D"
5489. .5608
/rpt_family="MIR"
6091. .6253
/rpt_family="L2"
complement(8466. .8772)
/rpt_family="AluSx"
8808. .8855
/rpt_family="L2"
9025. .9104
/rpt_family="MADE1"
complement(9180. .9204)
/rpt_family="(TA)n"
complement(9259. .9630)
/rpt_family="L1MA4A"
complement(10007. .10245)
/rpt_family="LTR16C"
10982. .11738
/rpt_family="L2"
complement(12270. .12301)
/rpt_family="(CA)n"
13179. .13285
/rpt_family="MER5A"
complement(13353. .13524)
/rpt_family="MER5B"
13562. .13588
/rpt_family="(TAA)n"
15505. .15567
/rpt_family="AT-rich"
complement(15707. .15940)
/rpt_family="AluSq"
complement(16926. .19866)
/rpt_family="L1PA3"
17507. .17684
/standard_name="G02335"
19864. .20661
/db_xref="dbSTS:7821"
/rpt_family="L1PA3"
complement(20839. .20909)
/rpt_family="L2"
complement(22349. .22640)
/rpt_family="AluSq"
complement(22769. .23927)
/rpt_family="L1MB5"
complement(23918. .24010)
/rpt_family="L1M4"
24012. .24322
/rpt_family="AluSx"
complement(24323. .24974)
/rpt_family="L1M4"
24975. .25052
/rpt_family="L1P4"
26902. .27010


```
repeat_region 2990..3195
/rpt_family="LIM4"
repeat_region 3385..3687
/rpt_family="AluSx"
repeat_region complement(4011..4177)
/rpt_family="AluJo"
repeat_region complement(4575..4877)
/rpt_family="AluSx"
repeat_region complement(5051..5176)
/rpt_family="FLAM_A"
repeat_region 5763..5926
/rpt_family="LIMB8"
repeat_region 5936..6065
/rpt_family="AluJo"
repeat_region 6066..6364
/rpt_family="AluJo"
repeat_region 6365..6534
/rpt_family="AluJB"
repeat_region 6535..6830
/rpt_family="LIMB8"
repeat_region 6831..7127
/rpt_family="AluSg"
repeat_region 7128..7761
/rpt_family="LIMB8"
repeat_region 8317..8477
/rpt_family="MER106A"
repeat_region 8983..9147
/rpt_family="AluSg/x"
repeat_region 10367..10391
/rpt_family="AluSg/x"
repeat_region 10541..10821
/rpt_family="(GA)n"
repeat_region 10823..11116
/rpt_family="AluJo"
repeat_region complement(13211..13451)
/rpt_family="AluJo"
repeat_region 18342..18487
/rpt_family="MIR"
repeat_region 18740..18818
/rpt_family="MIR"
repeat_region 18985..19240
/rpt_family="MT1F1"
repeat_region 19262..19395
/rpt_family="MT1F"
repeat_region 20786..21194
/rpt_family="MT2D"
repeat_region 21687..21872
/rpt_family="(CACGT)n"
repeat_region 21918..22100
/rpt_family="(CACGC)n"
repeat_region complement(24195..24499)
/rpt_family="AluSg"
repeat_region 24628..24906
/rpt_family="AluJo"
repeat_region 25099..25216
/rpt_family="AluSg"
repeat_region 25217..25518
/rpt_family="AluY"
repeat_region 25515..25697
/rpt_family="AluSg"
repeat_region 27078..27132
/rpt_family="MIR"
repeat_region 30600..30845
/rpt_family="AluJB"
repeat_region 31263..31336
/rpt_family="(TA)n"
repeat_region complement(31338..31634)
/rpt_family="AluSg"
repeat_region 31642..31916
/rpt_family="Tigger3 (Golem)"
repeat_region 32272..32572
/rpt_family="AluY"
repeat_region complement(35401..35707)

/rpt_family="AluSx"
complement(39516..39819)
/rpt_family="AluSx"
39820..40296
/rpt_family="LIMB7"
41019..41147
/rpt_family="FLAM_C"
42718..43016
/rpt_family="AluSg"
42843..42881
/note="single clone coverage"
42868..42877
/note="<30 qual single clone coverage"

Query Match 84.8%; Score 17.8; DB 9; Length 148509;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
|||||
Db 6568 TCATATATCCCAATAGAGACT 6588

RESULT 13
AC122451/c
LOCUS AC122451 160256 bp DNA linear ROD 05-NOV-2003
DEFINITION Mus musculus BAC clone RP24-252G15 from 16, complete sequence.
ACCESSION AC122451
VERSION AC122451.3 GI:21844638
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Shah,N. and Meyer,R.
TITLE 1 (bases 1 to 160256)
JOURNAL Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 2 (bases 1 to 160256)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 160256)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 160256)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 160256)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (16-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 160256)
AUTHORS Wilson,R.
TITLE Direct Submision
JOURNAL Submitted (05-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 16, 2002 this sequence version replaced gi:21490596.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BB0252G15
-----

NOTICE: This sequence may not represent the entire insert of this
```


/db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-409N18"
 /clone_lib="RPC1-23"

ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 176159;
 Best Local Similarity 90.5%; Pred. No. 1.2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
 ||| |||||
 Db 98074 TCATTAATCCCAATTAGAGACT 98054

Search completed: May 7, 2004, 10:43:39
 Job time : 673.597 secs

CC present invention are useful for diagnosing and/or prognosing disorders
 CC associated with an aberrant inflammatory response such as asthma,
 CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
 CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
 CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
 CC disease, atopic and contact dermatitis. The nucleic acid molecules can
 CC also be useful for identifying an individual amongst other individuals
 CC from the same species for use in forensic medicine and paternity testing.
 CC This polynucleotide sequence represents DNA relating to the human 5-
 CC lipoxigenase (5-LO) gene of the invention
 XX
 SQ Sequence 21 BP; 6 A; 4 C; 3 G; 7 T; 0 U; 1 Other;

Query Match 98.1%; Score 20.6; DB 6; Length 21;
 Best Local Similarity 95.2%; Pred. No. 1.5;
 Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
 |||||
 DB 1 TCATGTATCCCAATTAGAGACT 21
 |||||

RESULT 2
 ID ABT11118 standard; DNA; 21 BP.
 AC ABT11118;
 XX
 XX
 DT 05-DEC-2002 (first entry)
 XX
 XX Human 5-lipoxygenase gene related DNA sequence SEQ ID No 6.

XX Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
 KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
 KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
 KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
 KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
 KW ds.

XX Homo sapiens.
 XX WO200262825-A2.
 XX
 XX 15-AUG-2002.
 XX
 XX 07-FEB-2002; 2002WO-US003546.
 XX
 XX 08-FEB-2001; 2001US-0267515P.
 XX 21-AUG-2001; 2001US-0314248P.
 XX (MILL-) MILLENNIUM PHARM INC.

PI Barnes G, Meyer J;
 XX
 XX WPI; 2002-627522/67.
 XX
 XX New isolated nucleic acid molecule with an allelic variant of a
 PT polymorphic region of an 5-LO gene, useful for diagnosing and/or
 PT prognosticating disorders associated with an aberrant inflammatory
 PT response such as asthma.

XX Claim 1; Page 235; 290pp; English.
 XX
 XX The invention relates to an isolated human nucleic acid molecule
 CC comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
 CC (5-LO) gene, where the allelic variant comprises one or more nucleotide
 CC selected from any of 3, 20 or 21 base pair sequences, given in the
 CC specification, or their complement. The compositions and methods of the
 CC present invention are useful for diagnosing and/or prognosing disorders
 CC associated with an aberrant inflammatory response such as asthma,
 CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
 CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
 CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory

CC disease, atopic and contact dermatitis. The nucleic acid molecules can
 CC also be useful for identifying an individual amongst other individuals
 CC from the same species for use in forensic medicine and paternity testing.
 CC This polynucleotide sequence represents DNA relating to the human 5-
 CC lipoxigenase (5-LO) gene of the invention
 XX
 SQ Sequence 21 BP; 6 A; 4 C; 4 G; 7 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 6; Length 21;
 Best Local Similarity 95.2%; Pred. No. 5.9;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
 |||||
 DB 1 TCATGTATCCCAATTAGAGACT 21
 |||||

RESULT 3
 ID AAT88433 standard; DNA; 2177 BP.
 XX
 AC AAT88433;
 XX
 DT 14-MAY-1998 (first entry)
 XX
 XX Human 5-lipoxygenase gene polymorphism 1669 to 1680 deletion.
 DE
 DE Inflammatory disease; polymorphism; 5-lipoxygenase; asthma;
 KW ulcerative colitis; bronchitis; sinusitis; psoriasis; rhinitis;
 KW arthritis; diagnosis; treatment; ds.

XX Homo sapiens.
 OS Synthetic.
 XX
 XX WO9742347-A2.
 XX
 XX 13-NOV-1997.

XX 29-APR-1997; 97WO-US007137.
 XX
 XX 06-MAY-1996; 96US-0016890P.
 PR 25-APR-1997; 97US-00846020.
 XX
 XX (BGM) BRIGHAM & WOMENS HOSPITAL.
 XX
 XX Drazen JM, In K, Asano K, Beier D, Grobholz J;
 XX
 XX WPI; 1997-558997/51.

XX Classifying patients with inflammatory disease, specifically asthma -
 PT according to polymorphisms in 5-lipoxygenase gene regulatory region, e.g.
 PT to identify candidates for lipoxygenase inhibitor treatment.

XX Claim 13; Page; 56pp; English.

XX The present sequence was used in the development of a novel method for
 CC classifying patients suffering from an inflammatory disease. The method
 CC comprises identifying in DNA from at least 1 patient a sequence
 CC polymorphism, as compared with the normal 5-lipoxygenase (5-LOX) gene
 CC (AAT88431), in a 5-LOX regulatory gene sequence. The method can be
 CC applied to subjects with asthma, ulcerative colitis, bronchitis,
 CC sinusitis, psoriasis, allergic and non-allergic rhinitis, lupus or
 CC rheumatoid arthritis. Specifically it can be used to diagnose asthma or
 CC susceptibility to disease, identify treatments suitable for individual
 CC patients or assess the likely success of treatment. N.B. Sequence not
 CC given in the specification, but constructed using the wild type human 5-
 CC lipoxigenase gene sequence given on pages 40 to 41

SQ Sequence 2177 BP; 541 A; 597 C; 654 G; 385 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 2; Length 2177;
 Best Local Similarity 95.2%; Pred. No. 8.1;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCATGATCCAAATTAGAGACT 21
 Db 549 TCATGATCCGATTAGAGACT 569
 RESULT 4
 ID AAT88432 standard; DNA; 2183 BP.
 AC AAT88432;
 XX
 DT 14-MAY-1998 (first entry)
 DE Human 5-lipoxygenase gene polymorphism 1669 to 1674 deletion.
 XX
 KW Inflammatory disease; polymorphism; 5-lipoxygenase; asthma;
 KW ulcerative colitis; bronchitis; sinusitis; psoriasis; rhinitis;
 KW arthritis; diagnosis; treatment; ds.
 OS Homo sapiens.
 OS Synthetic.
 XX WO9742347-A2.
 XX 13-NOV-1997.
 XX 29-APR-1997; 97WO-US0071137.
 XX 06-MAY-1996; 96US-0016890P.
 XX 25-APR-1997; 97US-00846020.
 XX (BGHM) BRIGHAM & WOMENS HOSPITAL.
 PI Drazen JM, In K, Asano K, Beier D, Grobholz J;
 DR WPI; 1997-558997/51.
 XX
 PT Classifying patients with inflammatory disease, specifically asthma -
 PT according to polymorphisms in 5-lipoxygenase gene regulatory region, e.g.
 PT to identify candidates for lipoxygenase inhibitor treatment.
 XX
 PS Claim 12; Page; 56pp; English.
 XX
 CC The present sequence was used in the development of a novel method for
 CC classifying patients suffering from an inflammatory disease. The method
 CC comprises identifying in DNA from at least 1 patient a sequence
 CC polymorphism, as compared with the normal 5-lipoxygenase (5-LOX) gene
 CC (AAT88431), in a 5-LOX regulatory gene sequence. The method can be
 CC applied to subjects with asthma, ulcerative colitis, bronchitis,
 CC sinusitis, psoriasis, allergic and non-allergic rhinitis, lupus or
 CC rheumatoid arthritis. Specifically it can be used to diagnose asthma or
 CC susceptibility to disease, identify treatments suitable for individual
 CC patients or assess the likely success of treatment.
 XX
 PS Sequence 2183 BP; 541 A; 598 C; 659 G; 385 T; 0 U; 0 Other;
 XX
 CC Query Match 92.4%; Score 19.4; DB 2; Length 2183;
 CC Best Local Similarity 95.2%; Pred. No. 8.1;
 CC Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 TCATGATCCAAATTAGAGACT 21
 Db 549 TCATGATCCGATTAGAGACT 569
 RESULT 5
 ID AAT88431 standard; DNA; 2189 BP.
 AC AAT88431;
 XX
 DT 05-DEC-2002 (first entry)
 DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 1.
 XX
 KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
 KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
 KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
 KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
 KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
 KW ds.
 OS Homo sapiens.
 XX

XX 14-MAY-1998 (first entry)
 XX Human 5-lipoxygenase gene.
 XX
 KW Inflammatory disease; polymorphism; 5-lipoxygenase; asthma;
 KW ulcerative colitis; bronchitis; sinusitis; psoriasis; rhinitis;
 KW arthritis; diagnosis; treatment; ds.
 OS Homo sapiens.
 XX WO9742347-A2.
 XX 13-NOV-1997.
 XX 29-APR-1997; 97WO-US0071137.
 XX 06-MAY-1996; 96US-0016890P.
 XX 25-APR-1997; 97US-00846020.
 XX (BGHM) BRIGHAM & WOMENS HOSPITAL.
 PI Drazen JM, In K, Asano K, Beier D, Grobholz J;
 DR WPI; 1997-558997/51.
 XX
 PT Classifying patients with inflammatory disease, specifically asthma -
 PT according to polymorphisms in 5-lipoxygenase gene regulatory region, e.g.
 PT to identify candidates for lipoxygenase inhibitor treatment.
 XX
 PS Claim 11; Page 40-41; 56pp; English.
 XX
 CC The present sequence was used in the development of a novel method for
 CC classifying patients suffering from an inflammatory disease. The method
 CC comprises identifying in DNA from at least 1 patient a sequence
 CC polymorphism, as compared with the normal 5-lipoxygenase (5-LOX) gene
 CC (AAT88431), in a 5-LOX regulatory gene sequence. The method can be
 CC applied to subjects with asthma, ulcerative colitis, bronchitis,
 CC sinusitis, psoriasis, allergic and non-allergic rhinitis, lupus or
 CC rheumatoid arthritis. Specifically it can be used to diagnose asthma or
 CC susceptibility to disease, identify treatments suitable for individual
 CC patients or assess the likely success of treatment.
 XX
 PS Sequence 2189 BP; 541 A; 599 C; 664 G; 385 T; 0 U; 0 Other;
 XX
 CC Query Match 92.4%; Score 19.4; DB 2; Length 2189;
 CC Best Local Similarity 95.2%; Pred. No. 8.1;
 CC Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 TCATGATCCAAATTAGAGACT 21
 Db 549 TCATGATCCGATTAGAGACT 569
 RESULT 6
 ID AAT11113 standard; DNA; 2189 BP.
 XX AAT11113;
 XX
 DT 05-DEC-2002 (first entry)
 DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 1.
 XX
 KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
 KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
 KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
 KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
 KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
 KW ds.
 OS Homo sapiens.
 XX

comprising an allelic variant of a polymorphic region of a 5-lipoxygenase (5-LO) gene, where the allelic variant comprises one or more nucleotide selected from any of 3, 20 or 21 base pair sequences, given in the specification, or their complement. The compositions and methods of the present invention are useful for diagnosing and/or prognosing disorders associated with an aberrant inflammatory response such as asthma, bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis, rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory disease, atopic and contact dermatitis. The nucleic acid molecules can also be useful for identifying an individual amongst other individuals from the same species for use in forensic medicine and paternity testing. This polynucleotide sequence represents DNA relating to the human 5-lipoxygenase (5-LO) gene of the invention

Sequence 168273 BP; 46834 A; 36467 C; 36966 G; 46498 T; 0 U; 1508 Other;
 Query Match 92.4%; Score 19.4; DB 6; Length 168273;
 Best Local Similarity 95.2%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGATCCAAATTAGAGACT 21
 Db 167017 TCATGATCCGATTAGAGACT 167037

RESULT 11
 ADA02930
 ID ADA02930 standard; DNA; 96596 BP.

AC ADA02930;

DT 06-NOV-2003 (first entry)

DE Human PIK3R1 carcinoma associated gene, SEQ ID NO:1448.

Human; carcinoma associated; oncogene; carcinoma; cancer; breast; prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening; gene; ds.

OS Homo sapiens.

PN WO2003057146-A2.

PD 17-JUL-2003.

PF 26-DEC-2002; 2002WO-US041414.

PR 26-DEC-2001; 2001US-00035832.

PA (SAGR-) SAGRES DISCOVERY.

PI Morris DW;

DR WPI; 2003-587068/55.

PT New recombinant nucleic acid encoding carcinoma associated protein, useful for preparing compositions for treating carcinomas.

PS Claim 1; SEQ ID NO 1448; 245pp; English.

The invention relates to recombinant carcinoma associated (CA) nucleic acid sequences from mouse and human (ADA01482-ADA03094), and to recombinant carcinoma associated proteins (CAP) encoded by them. The invention also encompasses expression vectors and host cells comprising a CA nucleic acid, a polypeptide (especially an antibody) that specifically binds to the protein, and a biochip comprising CA nucleic acid or fragments thereof. The sequences of the invention were identified using oncogenic retroviruses, which insert into the genome of the host organism at random. Many of these do not carry transduced host oncogenes or pathogenic trans-acting viral genes, meaning that cancer incidence is a direct consequence of the effects of proviral integration into host protooncogenes. The CA nucleic acid sequences can be used to diagnose

carcinoma (especially breast cancer, prostate cancer, lymphoma or leukaemia) or a propensity to carcinoma by determination of the sequence of a CA gene, or by determination of CA gene expression in particular tissues. CA nucleic acids, proteins and antibodies are also useful as therapeutic agents and in screening and evaluating drug candidates. The present sequence represents a specifically claimed human CA nucleic acid sequence of the invention. Note: The complete sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 8; Length 96596;

Best Local Similarity 90.0%; Pred. No. 2.1e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGTATCCAAATTAGAGACT 21

Db 36245 CATGTATCAATCAGAGACT 36264

RESULT 12

ADB72668

ID ADB72668 standard; DNA; 96596 BP.

AC ADB72668;

DT 04-DEC-2003 (first entry)

DE Human PIK3R1 gene.

human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas; cancer; neoplasm; adenocarcinoma; sarcoma; gene.

OS Homo sapiens.

PN WO2003008583-A2.

PD 30-JAN-2003.

PF 26-DEC-2001; 2001WO-US051291.

PR 02-MAR-2001; 2001US-00798586.

PR 23-OCT-2001; 2001US-00004113.

PR 08-NOV-2001; 2001US-00052482.

PR 30-NOV-2001; 2001US-00997722.

PR 20-DEC-2001; 2001US-00034650.

PA (SAGR-) SAGRES DISCOVERY.

PI Morris DW, Engelhard EK;

DR WPI; 2003-239337/23.

PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas, cancers, neoplasm, adenocarcinoma, or sarcomas.

PS Claim 1; SEQ ID NO 496; 2304pp; English.

The invention relates to a novel recombinant nucleic acid comprising a nucleotide sequence selected from any of the 660 sequences fully defined in the specification. A polynucleotide of the invention has cytostatic activity, and may have a use in gene therapy, or in a vaccine. The recombinant nucleic acids and polypeptides are useful for treating carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and sarcomas. The present sequence represents a human gene of the invention.

Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

Query Match

Best Local Similarity 80.0%; Score 16.8; DB 9; Length 96596;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      2 CATGTATCCAATTAGAGACT 21
Db      36245 CATGTATCCAATCAGAGACT 36264

RESULT 13
ADC85410
ID      ADC85410 standard; DNA; 96596 BP.
XX
AC      ADC85410;
XX
DT      01-JAN-2004 (first entry)
XX
DE      Human Pik3r1 genomic sequence.
XX
KW      Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW      secreted; transmembrane; intracellular; ds.
XX
OS      Homo sapiens.
XX
PN      WO2003045230-A2.
XX
PD      05-JUN-2003.
XX
PF      02-DEC-2002; 2002WO-US038582.
XX
PR      30-NOV-2001; 2001US-00997722.
XX
PA      (SAGR-) SAGRES DISCOVERY.
XX
PI      Morris DW, Engelhard EK;
DR      WPI; 2003-513603/48.
XX
PT      New recombinant nucleic acid comprising a nucleotide sequence of any of
PT      the carcinoma-associated (CA) genes, useful for screening for drug
PT      candidates for diagnosing or treating carcinomas.
XX
PS      Claim 1; SEQ ID NO 196; 983pp; English.
XX
CC      The invention relates to a recombinant nucleic acid comprising a
CC      nucleotide sequence selected from any of the fully defined carcinoma-
CC      associated (CA) genes from the 50 tables given in the specification. The
CC      CA proteins are secreted, transmembrane or intracellular proteins. The
CC      recombinant nucleic acids are useful for screening for drug candidates
CC      for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC      ADC85514 represent CA genes of the invention.
XX
SQ      Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

Query Match      80.0%; Score 16.8; DB 9; Length 96596;
Best Local Similarity 90.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CATGTATCCAATTAGAGACT 21
Db      36245 CATGTATCCAATCAGAGACT 36264

RESULT 14
AAH76381
ID      AAH76381 standard; DNA; 1660 BP.
XX
AC      AAH76381;
XX
DT      30-NOV-2001 (first entry)
XX
DE      O. vulgaris tachykinin-related peptide encoding DNA.
XX
KW      Tachykinin-related peptide; drug; agricultural; tachykinin;
KW      neurotransmission; ds.
XX

```

```

OS      Octopus vulgaris.
XX
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      384..1181
FT      /*tag= a
XX
PN      JP2001151796-A.
XX
PD      05-JUN-2001.
XX
PF      11-MAY-2000; 2000JP-00138902.
XX
PR      14-SEP-1999; 99JP-00260982.
XX
PA      (SUNR ) SUNTORY LTD.
XX
WPI; 2001-560300/63.
DR      P-PSDB; AAB85881.
XX
PT      A new tachykinin-related peptide (I), its precursor and a gene for
PT      encoding (I), is useful as a biochemical reagent for elucidating
PT      neurotransmission system.
XX
PS      Claim 7; Fig 10; 19pp; Japanese.
XX
CC      The invention provides a tachykinin-related peptide of the formula, R1-
CC      Phe-Xaa-Yaa-Zaa-Arg-NH 2 where R1 = a peptide of the amino acid residue
CC      Nos. of 5 or 6; Xaa = Leu, Ile, Met, Val or Gln; Yaa = Gly or Pro; and
CC      Zaa = Ser or Thr. The tachykinin-related peptide can be expressed by
CC      standard recombinant methodology. It is useful as an active component in
CC      a drug or an agricultural chemical. The peptide is useful as a
CC      biochemical reagent for elucidating neurotransmission system. The present
CC      sequence represents a DNA encoding the tachykinin-related peptide from
CC      Octopus vulgaris
XX
SQ      Sequence 1660 BP; 546 A; 360 C; 264 G; 490 T; 0 U; 0 Other;

Query Match      78.1%; Score 16.4; DB 4; Length 1660;
Best Local Similarity 94.4%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCATGTATCCAATTAGAG 18
Db      382 TCATGTATGCAATTAGAG 399
      ||||| ||||| ||||| |||||
      ||||| ||||| ||||| |||||

RESULT 15
AAC04867/c
ID      AAC04867 standard; cDNA; 394 BP.
XX
AC      AAC04867;
XX
DT      06-OCT-2000 (first entry)
XX
DE      Human secreted protein 5' EST, SEQ ID NO: 8942.
XX
KW      Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW      gene therapy; chromosome mapping; ss.
XX
OS      Homo sapiens.
XX
PN      EF1033401-A2.
XX
PD      06-SEP-2000.
XX
PF      21-FEB-2000; 2000EP-00200610.
XX
PR      26-FEB-1999; 99US-0122487P.
XX
PA      (GEST ) GENSET.
XX
PI      Dumas Milne Edwards J, Duclert A, Giordano J;
XX

```

DR WPI; 2000-500381/45.
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX
 XX Claim 1; SEQ ID NO 8942; 71pp + Sequence Listing; English.
 PS
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors
 XX
 SQ Sequence 394 BP; 89 A; 79 C; 68 G; 145 T; 0 U; 13 Other;
 Query Match 77.1%; Score 16.2; DB 3; Length 394;
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;
 Matches 15; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TCATGTATCCATTAGACT 21
 Db 119 TCATATATCCCATWRRGACT 99

Search completed: May 7, 2004, 10:10:30
 Job time : 152.048 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 11:54:38 ; Search time 148.016 Seconds
(without alignments)
642.699 Million cell updates/sec

Title: US-10-071-411A-6

Perfect score: 21

Sequence: 1 tcattgtatccaattagagact 21

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	2189	15	US-10-240-305-5
2	17.4	82.9	254087	13	US-10-087-192-223
3	16.8	80.0	555	13	US-10-027-632-231239
4	16.8	80.0	555	16	US-10-027-632-231239
5	16.8	80.0	735	13	US-10-027-632-21789
6	16.8	80.0	735	16	US-10-027-632-21789
7	16.8	80.0	96596	12	US-09-997-722-196
8	16.4	78.1	676	13	US-10-027-632-231871
9	16.4	78.1	676	16	US-10-027-632-231871
10	16.2	77.1	331	13	US-10-085-783A-12549
11	16.2	77.1	331	16	US-10-085-783A-12549
12	16.2	77.1	528	13	US-10-027-632-181959
13	16.2	77.1	528	16	US-10-027-632-181959
14	16.2	77.1	637	13	US-10-027-632-233675

15	16.2	77.1	637	16	US-10-027-632-233675
16	16.2	77.1	2220	13	US-10-027-632-103330
17	16.2	77.1	2220	16	US-10-027-632-103330
18	16.2	77.1	2431	13	US-10-027-632-250429
19	16.2	77.1	2431	16	US-10-027-632-250429
20	16.2	77.1	2431	13	US-10-027-632-250429
21	16.2	77.1	2431	16	US-10-027-632-250429
22	16.2	77.1	49888	16	US-10-085-117-25
23	16.2	77.1	72604	15	US-10-162-497-7
24	16.2	77.1	295096	13	US-10-087-192-331
25	16.2	77.1	653122	13	US-10-087-192-226
26	16.2	77.1	1503841	9	US-09-795-668-1
27	16.2	77.1	1503841	9	US-09-795-668-1
28	16.2	77.1	1503841	9	US-09-946-807-1
29	16.2	77.1	3186778	13	US-10-027-632-174961
30	16.2	77.1	3186778	16	US-10-027-632-174961
31	15.8	75.2	533	13	US-10-027-632-323061
32	15.8	75.2	533	16	US-10-027-632-323061
33	15.8	75.2	616	13	US-10-027-632-230612
34	15.8	75.2	616	16	US-10-027-632-230612
35	15.8	75.2	637	13	US-10-027-632-218904
36	15.8	75.2	637	16	US-10-027-632-218904
37	15.8	75.2	637	13	US-10-027-632-218905
38	15.8	75.2	637	16	US-10-027-632-218905
39	15.8	75.2	637	16	US-10-027-632-218904
40	15.8	75.2	637	16	US-10-027-632-218905
41	15.8	75.2	964	13	US-10-425-114-11009
42	15.8	75.2	1183	13	US-10-424-599-48302
43	15.8	75.2	1228	13	US-10-425-114-5943
44	15.8	75.2	1270	13	US-10-424-599-47488
45	15.8	75.2	1412	9	US-09-938-842A-3608

ALIGNMENTS

RESULT 1

US-10-240-305-5
; Sequence 5, Application US/10240305
; Publication No. US20030162193A1
; GENERAL INFORMATION:
; APPLICANT: GLAXO GROUP LIMITED
; APPLICANT: EDWARDS, WAYNE H.
; APPLICANT: EDWARDS, Lisa D.
; APPLICANT: EMMETT, Amanda H.
; APPLICANT: PILLAI, Sreekumar
; APPLICANT: SPRANKEL, Catherine S.
; TITLE OF INVENTION: Medicine Response Assay in Respiratory Disease
; FILE REFERENCE: PU3958 & PU4254
; CURRENT APPLICATION NUMBER: US/10/240,305
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-240-305-5

Query Match 92.4%; Score 19.4; DB 15; Length 2189;
Best Local Similarity 95.2%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGTCATCCAAATTAGAGACT 21
|||||
Db 549 TCATGTCATCCAAATTAGAGACT 569

RESULT 2

US-10-087-192-223/c
; Sequence 223, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:

APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 223
LENGTH: 254087
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(254087)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-223

Query Match 82.9%; Score 17.4; DB 13; Length 254087;
Best Local Similarity 94.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGTATCCAATTAGAC 20
DB 181253 CATGTATCCAATTAGTAC 181235

RESULT 3
US-10-027-632-231239
Sequence 231239, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 231239
LENGTH: 555
TYPE: DNA
ORGANISM: Human
US-10-027-632-231239

Query Match 80.0%; Score 16.8; DB 13; Length 555;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGTATCCAATTAGAC 21
DB 93 CATGTATCCAATCAGAGACT 112

RESULT 4
US-10-027-632-231239
Sequence 231239, Application US/10027632
Publication No. US20020204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 231239
LENGTH: 555
TYPE: DNA
ORGANISM: Human
US-10-027-632-231239

Query Match 80.0%; Score 16.8; DB 16; Length 555;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGTATCCAATTAGAGACT 21
DB 93 CATGTATCCAATCAGAGACT 112

RESULT 5
US-10-027-632-21789/c
Sequence 21789, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21789
LENGTH: 735
TYPE: DNA

;
; ORGANISM: Human
US-10-027-632-21789

Query Match 80.0%; Score 16.8; DB 13; Length 735;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTATCCAATTAGAGAC 20
|||||
Db 459 TCATTATCCAATTGGAGAC 440

RESULT 6

US-10-027-632-21789/c
; Sequence 21789, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21789
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-21789

Query Match 80.0%; Score 16.8; DB 16; Length 735;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTATCCAATTAGAGAC 20
|||||
Db 459 TCATTATCCAATTGGAGAC 440

RESULT 7

US-09-997-722-196
; Sequence 196, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 196
; LENGTH: 96596

;
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-997-722-196

Query Match 80.0%; Score 16.8; DB 12; Length 96596;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATGTATCCAATTAGAGACT 21
|||||
Db 36245 CATGTATCCAATTAGAGACT 36264

RESULT 8

US-10-027-632-231871
; Sequence 231871, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231871
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-231871

Query Match 78.1%; Score 16.4; DB 13; Length 676;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGTATCCAATTAGAG 18
|||||
Db 403 TCATGAATCCAATTAGAG 420

RESULT 9

US-10-027-632-231871
; Sequence 231871, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218

```
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/271,955
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231871
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-231871

Query Match
Best Local Similarity 78.1%; Score 16.4; DB 16; Length 676;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAG 18
DB 403 TCATGAATCCCAATTAGAG 420

RESULT 10
US-10-085-783A-12549/c
; Sequence 12549, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12549
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (137)..(137)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-12549

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 13; Length 331;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
DB 104 TCATGTATCCGATAAGGACT 84

RESULT 11
US-10-242-535A-12549/c
; Sequence 12549, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
```

```
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12549
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (137)..(137)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-12549

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 16; Length 331;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
DB 104 TCATGTATCCGATAAGGACT 84

RESULT 12
US-10-027-632-181959
; Sequence 181959, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181959
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-181959

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 13; Length 528;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
DB 112 TCATGTATCCCAATTAAACT 132

RESULT 13
US-10-027-632-181959
```



```
; Sequence 181959, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181959
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-181959
```

```
Query Match      77.1%; Score 16.2; DB 16; Length 528;
Best Local Similarity 85.7%; Pred. No. 4.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1 TCATGTATCCCAATTAGAGACT 21
      ||||| ||||| ||||| |||||
Db      112 TCATGTAGCCCAATTAAAAACT 132
```

```
RESULT 14
US-10-027-632-233675
; Sequence 233675, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 233675
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-233675
```

```
Query Match      77.1%; Score 16.2; DB 13; Length 637;
Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 TCATGTATCCCAATTAGAGACT 21
      ||||| ||||| ||||| |||||
Db      101 TCATGTATCTGATAAGAGACT 121

RESULT 15
US-10-027-632-233675
; Sequence 233675, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 233675
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-233675
```

```
Query Match      77.1%; Score 16.2; DB 16; Length 637;
Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1 TCATGTATCCCAATTAGAGACT 21
      ||||| ||||| ||||| |||||
Db      101 TCATGTATCTGATAAGAGACT 121
```

```
Search completed: May 7, 2004, 13:42:30
Job time : 155.016 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 09:56:42 ; Search time 1438.5 Seconds
(without alignments)
435.944 Million cell updates/sec

Title: US-10-071-411A-6

Perfect score: 21

Sequence: 1 tcatgtatccaattagagact 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hcc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.4	87.6	667	28	AZ710684 RPCI-24-1
C 2	18	85.7	505	28	AQ227780 HS_2012B
C 3	17.8	84.8	120	9	AI137701 UI-R-C0-1
C 4	17.8	84.8	287	10	BB717266 BB717266

C 5	17.8	84.8	401	28	BH304012
C 6	17.8	84.8	584	28	AQ475307
C 7	17.8	84.8	671	28	BH279572
C 8	17.8	84.8	685	29	AG183795
C 9	17.8	84.8	693	28	BZ644002
C 10	17.8	84.8	735	28	BZ322372
C 11	17.8	84.8	768	28	BZ315856
C 12	17.8	84.8	847	28	BH112961
C 13	17.8	84.8	879	28	CC368094
C 14	17.4	81.0	851	28	BZ212281
C 15	17	81.0	328	28	B58442
C 16	16.8	80.0	221	12	BI140461
C 17	16.8	80.0	255	28	AQ094931
C 18	16.8	80.0	264	10	B5564270
C 19	16.8	80.0	266	12	BI140075
C 20	16.8	80.0	282	14	CA908115
C 21	16.8	80.0	333	14	CA638809
C 22	16.8	80.0	343	12	BI448394
C 23	16.8	80.0	352	9	AV311939
C 24	16.8	80.0	362	29	CE759815
C 25	16.8	80.0	389	29	CE384792
C 26	16.8	80.0	422	12	BJ033905
C 27	16.8	80.0	429	14	CB174581
C 28	16.8	80.0	438	9	AV590406
C 29	16.8	80.0	447	14	CD927118
C 30	16.8	80.0	459	9	AV597487
C 31	16.8	80.0	463	13	BU794223
C 32	16.8	80.0	483	9	AV597778
C 33	16.8	80.0	484	28	BH188998
C 34	16.8	80.0	484	29	CNS07SV2
C 35	16.8	80.0	496	29	CC819936
C 36	16.8	80.0	497	10	BF079691
C 37	16.8	80.0	513	28	AQ862367
C 38	16.8	80.0	514	10	BF079687
C 39	16.8	80.0	520	28	AQ971737
C 40	16.8	80.0	522	10	AW294697
C 41	16.8	80.0	522	12	BI140452
C 42	16.8	80.0	524	12	BI140066
C 43	16.8	80.0	544	28	AZ122299
C 44	16.8	80.0	546	13	BQ801908
C 45	16.8	80.0	546	14	CD802345

ALIGNMENTS

RESULT 1
AZ710684/c
LOCUS
DEFINITION RPCI-24-119D9.TV RPCI-24 Mus musculus genomic clone RPCI-24-119D9, linear GSS 24-JAN-2001
Genomic survey sequence.
ACCESSION AZ710684
VERSION AZ710684.1 GI:12442500
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 667)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shvartsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-119D9.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC

BH304012 CH230-98A
AQ475307 CITBI-E1-
BH279572 CH230-97B
AG183795 Pan trogl
BZ644002 OGAOI25TM
BZ322372 ia64h08.g
BZ315856 ia64h08.b
BH112961 RPCI-24-2
CC368094 PUDX65TD
BZ212281 CH230-331
B58442 CIT-HSP-201
BI140461 IPI 49 F1
AQ094931 GSSTC0808
B5564270 BB564270
BI140075 IPI 49 F1
CA908115 PCS03965
CA638809 wreln.pk0
BI448394 dac74f09.
AV311939 AV311939
CE759815 tigr-GSS-
CE384792 tigr-888-
BJ033905 BJ033905
CB174581 Pha-EST00
AV590406 AV590406
CD927118 GR45.100P
AV597487 AV597487
BU794223 SJF2CDC09
AV597778 AV597778
BH188998 037 A_08-
AL625937 T7 end of
CC819936 100006L14
BF079691 230362 MA
AQ862367 nbab0018D
BF079687 230356 MA
AQ971737 RPCI-23-3
AW294697 UI-R-BW0-
BI140452 IPI 49 E1
BI140066 IPI 49 E1
AZ122299 RPCI-23-3
BQ801908 WHE2819.H
CD802345 EST 9825

library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choori.org/bacpac/orderingframe.html>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 119 row: D column: 9
 Seq primer: T7
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..567
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-119D9"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /clone_lib="RPCI-24"
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

ORIGIN

Query Match 87.6%; Score 18.4; DB 28; Length 667;
 Best Local Similarity 95.0%; Pred. No. 2.8e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGATCCCAATTAGAGC 20
 |||
 Db 634 TCATGATCCCAATTAGAGC 615

RESULT 2

AQ227780/c
 LOCUS AQ227780 505 bp DNA linear GSS 26-SEP-1998
 DEFINITION HS_2012 B2 G10 MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2012 Col=20 Row=N, genomic survey sequence.

ACCESSION AQ227780.1 GI:3653009

VERSION GSS.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 505)

REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

10449764

CONTACT: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2012 row: N column: 20

Class: BAC ends

High quality sequence stop: 505.

Location/Qualifiers

1..505

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="Plate=2012 Col=20 Row=N"

/sex="male"

ORIGIN

Query Match 85.7%; Score 18; DB 28; Length 505;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ATGTATCCAATTAGAGC 20

|||

Db 322 ATGTATCCAATTAGAGC 305

RESULT 3

AI137701

LOCUS AI137701 120 bp mRNA linear EST 05-JUL-1999
 DEFINITION UI-R-CO-ig-e-06-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone UI-R-CO-ig-e-06-0-UI 3', mRNA sequence.

ACCESSION AI137701

VERSION AI137701.1

KEYWORDS GI:3638478

SOURCE EST.

ORGANISM Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 120)

REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

CONTACT: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Brain library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE ID=1782264

Seq primer: M13 Forward

POLYA=No.

FEATURES

Location/Qualifiers

1..120

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CO-ig-e-06-0-UI"

/dev_stages="adult"

/lab_host="DHI0B (Life Technologies)"

/clone_lib="UI-R-CO"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CO library is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-CO) was constructed as follows: PCR amplified cDNA inserts from a

pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonardo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 120;
Best Local Similarity 90.5%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
|||||
Db 89 TCATGTATCCCAAGTAGAGAGT 109

RESULT 4

BB717266 287 bp mRNA linear EST 11-OCT-2001
LOCUS BB717266 RIKEN full-length enriched, adult male liver tumor Mus
DEFINITION musculus cDNA clone C730005B08 3', mRNA sequence.

BB717266 1 GI:16070435
VERSION BB717266
KEYWORDS EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 287)

Akimura,T., Arawaka,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hirooka,T., Hiroane,T., Imctani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

Unpublished (2001)

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES

source

Location/Qualifiers
1..287
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="C730005B08"
/sex="male"
/tissue_type="liver tumor"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male liver
tumor"

Notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCACTCGAGTTTTTTTTTTTNN 3'], cDNA was
GAGAGAGAGCGCGCACTCGAGTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATAATTAATCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC 1.
Tissue was provided by William A. Held, Roswell Park
Cancer Institute, Department of Molecular and Cellular
Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose
assistance we gratefully acknowledge."

ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 287;
Best Local Similarity 90.5%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21

|||||
Db 132 TCATCTACCAATTAGAGACT 152

RESULT 5

BH304012/c

LOCUS

DEFINITION BH304012 401 bp DNA linear GSS 30-NOV-2001
CH230-98A21-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
BH230-98A21, genomic survey sequence.

ACCESSION BH304012

VERSION BH304012.1 GI:17216420

KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 401)

Zhao,S., Shetty,J., Shatsman,S., Tsengaye,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

JOURNAL

COMMENT

Other_GSSs: CH230-98A21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.html). For BAC library availability, please contact Pieter de Jong (pdejong@mail.choi.org). Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/or_ering_information.htm). BAC end page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html

Plate: 98 row: A column: 21

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. .401
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-98A21"
/sex="female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 84.8%; Score 17.8; DE 28; Length 401;
Best Local Similarity 90.5%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCATGTATCCCAATTAGAGACT 21

Db 159 TCATTATCCCACTTAGAGACT 139

RESULT 6

AQ475307/c

LOCUS CITBI-E1-2609J16.TR CITBI-E1 Homo sapiens genomic clone 2609J16,
genomic survey sequence.

ACCESSION AQ475307

VERSION AQ475307.1 GI:4657426

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 594)

AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and
Venter, J.C.

TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready

Map Building

JOURNAL Unpublished (1997)

COMMENT Other GSSs: CITBI-E1-2609J16.TF

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeetgr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

Location/Qualifiers

1. .584
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2609J16"
/sex="male"
/cell_type="sperm"
/clone_lib="CITBI-E1"
/note="Vector: pBelobAC11; Site 1: EcoRI; Site 2: EcoRI;

ORIGIN

CalTech Human BAC Library D"

Query Match 84.8%; Score 17.8; DB 28; Length 584;

Best Local Similarity 90.5%; Pred. No. 5.3e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCATGTATCCCAATTAGAGACT 21

Db 499 TCATGTATCTAATAAGAGACT 479

RESULT 7

BH279572/c

LOCUS

DEFINITION BH279572 671 bp DNA linear GSS 30-NOV-2001
CH230-97B22-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
genomic survey sequence.

ACCESSION BH279572

VERSION BH279572.1 GI:17191974

KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 671)

AUTHORS

Zhao, S., Shetty, J., Shatsman, S., Tsengaye, G., Geer, K.,
Sivatsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, F., de Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other GSSs: CH230-97B22.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.html). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.choi.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/or_ering_information.htm). BAC end

page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html

Plate: 97 row: B column: 22

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. .671
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-97B22"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 84.8%; Score 17.8; DB 28; Length 671;

Best Local Similarity 90.5%; Pred. No. 5.4e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCATGTATCCCAATTAGAGACT 21

Db 417 TCATTATCCCACTTAGAGACT 397

RESULT 8

```

AG183795/c
LOCUS      AG183795      685 bp      DNA      linear      GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-057G05.T7, genomic survey
sequence.
ACCESSION  AG183795
VERSION     AG183795.1  GI:16713475
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
REFERENCE  1
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      BAC end sequences of Library RPCI-43
JOURNAL    Unpublished
REFERENCE  2
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail: chimpansegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
            Tel: 81-45-503-9111, Fax: 81-45-503-9170)
            Clones are derived from the chimpanzee BAC library RPCI-43. This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
COMMENT    PRIMERS
LIBRARY    Sequencing: T7
            Vector : pBACe3.6
            R.Site 1 : EcoRI
            R.Site 2 : EcoRI
FEATURES   source
            Location/Qualifiers
                1..685
                /organism="Pan troglodytes"
                /mol_type="genomic DNA"
                /db_xref="taxon:9598"
                /clone="RP43-057G05.T7"
                /sex="male"
                /cell_type="lymphocytes"
                /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN
Query Match      84.8%; Score 17.8; DB 29; Length 685;
Best Local Similarity 90.5%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCATGTCATCAATTAGAGACT 21
        |||||
        515 TCATGTCATCAATTAGAGACT 495

Db

RESULT 9
BZ644002
LOCUS      OGAOI25TM ZM.0.7.1.5 KB Zea mays genomic clone ZMMBMA0108E01,
DEFINITION genomic survey sequence.
ACCESSION  BZ644002
VERSION     BZ644002.1  GI:28106166
KEYWORDS   GSS.
SOURCE     Zea mays
ORGANISM   Zea mays
REFERENCE  1
AUTHORS    Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
            Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
            Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
TITLE      Consortium for Maize Genomics
JOURNAL    Unpublished (2002)

```

```

COMMENT      Other GSSs: OGAOI25TC
              Contact: Cathy Whitelaw
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301-838-5843
              Fax: 301-838-0208
              Email: whitelaw@tigr.org
              Seq primer: TG
              Class: sheared ends.
FEATURES     Location/Qualifiers
              1..693
              /organism="Zea mays"
              /mol_type="genomic DNA"
              /strain="B73"
              /db_xref="taxon:4577"
              /clone="ZMMBMA0108E01"
              /clone_lib="ZM 0.7 1.5 KB"
              /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
              methylation filtered genomic DNA library"
ORIGIN
Query Match      84.8%; Score 17.8; DB 28; Length 693;
Best Local Similarity 90.5%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCATGTCATCAATTAGAGACT 21
        |||||
        258 TCATGTCATCAATTAGAGACT 278

Db

RESULT 10
BZ322372/c
LOCUS      BZ322372      735 bp      DNA      linear      GSS 06-NOV-2002
DEFINITION ia64h08.g1 WGS-ZmaysF (JMI07 adapted methyl filtered) Zea mays
            genomic clone ia64h08 5', genomic survey sequence.
ACCESSION  BZ322372
VERSION     BZ322372.1  GI:24699739
KEYWORDS   GSS.
SOURCE     Zea mays
ORGANISM   Zea mays
REFERENCE  1
AUTHORS    Rabinowicz, P.D., O'Shaughnessy, A.L., Balijs, V., Dedhia, N.,
            Kazenburger, F., King, L., Miller, B., Muller, S., Nascento, L.,
            Zutavern, T., McCombie, W.R. and Martienssen, R.A.
            Genomic shotgun sequences from Zea mays (methyl-filtered)
            Unpublished (2002)
            Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mcombie@cshl.org
            Plate: ia64 row: h column: 08
            Seq primer: -21M13UnivRev
            Class: shotgun
            High quality sequence stop: 735.
FEATURES     Location/Qualifiers
              1..735
              /organism="Zea mays"
              /mol_type="genomic DNA"
              /cultivar="B73"
              /db_xref="taxon:4577"
              /clone="ia64h08"
              /lab_host="JMI07 or DHSa"
              /clone_lib="WGS-ZmaysF (JMI07 adapted methyl filtered)"
              /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
              The vector was digested with XbaI and one nucleotide was
              added by fill in the recessive 3' end. The genomic DNA
              was nebulized, end repaired, adaptor ligated and size

```

fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (.x/y reads in M13mp19, .b/g reads in pUC19). The same ligation was transformed in either JM107 or DH5a. "

ORIGIN

Query Match 84.8%; Score 17.8; DB 28; Length 735;
Best Local Similarity 90.5%; Pred. No. 5.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCATGTATCCCAATTAGAGACT 21

Db 698 TGATGTATCCCAATTAGAGATT 678

RESULT 11

BZ315856 768 bp DNA linear GSS 06-NOV-2002
LOCUS ia64h08.b1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
DEFINITION genomic clone ia64h08 5', genomic survey sequence.

ACCESSION BZ315856

VERSION BZ315856.1 GI:24685967

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 768)
Rabinowicz, P.D., O'Shaughnessy, A.L., Balijs, V., Dedhia, N.,
Kutzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, T., McCombie, W.R. and Martienssen, R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)

JOURNAL

Unpublished (2002)

COMMENT

Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: ia64 row: h column: 08

Seq primer: -21M13UnivFwd

Class: shotgun

High quality sequence stop: 768.

FEATURES

source

1..768

/organism="Zea mays"

/mol_type="genomic DNA"

/cultivar="B73"

/db_xref="taxon:4577"

/clone="ia64h08"

/lab_host="JM107 or DH5a"

/clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"

/note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;

The vector was digested with XbaI and one nucleotide was

added by fill in in the recessive 3' end. The genomic DNA

was nebulized, end repaired, adaptor ligated and size

fractionated using sephadex. The resulting fragments were

between 0.8 and 3 kb and were cloned into the vector

(.x/y reads in M13mp19, .b/g reads in pUC19). The same

ligation was transformed in either JM107 or DH5a. "

ORIGIN

Query Match 84.8%; Score 17.8; DB 28; Length 768;
Best Local Similarity 90.5%; Pred. No. 5.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCATGTATCCCAATTAGAGACT 21

Db 312 TGATGTATCCCAATTAGAGATT 332

RESULT 12

BH112961

LOCUS

DEFINITION

RPCI-24-252G15-TJ RPCI-24 Mus musculus genomic clone

BH112961

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Other GSSs: RPCI-24-252G15.TV

Mouse BAC End Sequences from Library RPCI-24

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 252 row: G column: 15

Seq primer: SP6

Class: BAC ends

FEATURES

source

1..847

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-252G15"

/sex="Male"

/cell_type="Spleen/Brain"

/clone_lib="RPCI-24"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

ORIGIN

Query Match

Best Local Similarity

Matches

19; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

Qy

1 TCATGTATCCCAATTAGAGACT 21

Db

633 TCATTATCCCAATTAGAGCCT 653

RESULT 13

CC368094

LOCUS

DEFINITION

PUBDX65TD ZM.0.6 1.0 KB Zea mays genomic clone ZMMBta230K10,

genomic survey sequence.

ACCESSION CC368094

VERSION CC368094.1 GI:30837494

KEYWORDS GSS.

SOURCE

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

BH112961 847 bp DNA linear GSS 19-JUL-2001
RPCI-24-252G15-TJ RPCI-24 Mus musculus genomic clone
DEFINITION RPCI-24-252G15, genomic survey sequence.

BH112961

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Other GSSs: RPCI-24-252G15.TV

Mouse BAC End Sequences from Library RPCI-24

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 252 row: G column: 15

Seq primer: SP6

Class: BAC ends

FEATURES

Location/Qualifiers

1..847

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-252G15"

/sex="Male"

/cell_type="Spleen/Brain"

/clone_lib="RPCI-24"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

ORIGIN

Query Match

Best Local Similarity

Matches

19; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

Qy

1 TCATGTATCCCAATTAGAGACT 21

Db

633 TCATTATCCCAATTAGAGCCT 653

RESULT 13

CC368094

LOCUS

DEFINITION

PUBDX65TD ZM.0.6 1.0 KB Zea mays genomic clone ZMMBta230K10,

genomic survey sequence.

ACCESSION CC368094

VERSION CC368094.1 GI:30837494

KEYWORDS GSS.

SOURCE

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

The invention relates to an isolated human nucleic acid molecule comprising an allelic variant of a polymorphic region of a 5-lipoxygenase (5-LO) gene, where the allelic variant comprises one or more nucleotide selected from any of 3, 20 or 21 base pair sequences, given in the specification, or their complement. The compositions and methods of the

CC present invention are useful for diagnosing and/or prognosing disorders
CC associated with an aberrant inflammatory response such as asthma,
CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
CC disease, atopic and contact dermatitis. The nucleic acid molecules can
CC also be useful for identifying an individual amongst other individuals
CC from the same species for use in forensic medicine and paternity testing.
CC This polynucleotide sequence represents DNA relating to the human 5-
CC lipoxigenase (5-LO) gene of the invention
XX
SQ Sequence 21 BP; 5 A; 7 C; 5 G; 3 T; 0 U; 1 Other;

Query Match 98.1%; Score 20.6; DB 6; Length 21;
Best Local Similarity 95.2%; Pred. No. 0.47; Indels 0; Gaps 0;
Matches 20; Conservative 1; Mismatches 0;

Qy 1 AGGAGCGCGCAAAACCTTCTC 21
Db 1 AGGAGCGCGCAAAACCTTCTC 21

RESULT 2
ABT11116
ID ABT11116 standard; DNA; 21 BP.
XX
AC ABT11116;
XX
DT 05-DEC-2002 (first entry)
XX
DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 4.
XX
KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
KW ds.
XX
OS Homo sapiens.
XX
FN WO200262825-A2.
XX
PD 15-AUG-2002.
XX
PF 07-FEB-2002; 2002WO-US003546.
XX
PR 08-FEB-2001; 2001US-0267515P.
XX
PR 21-AUG-2001; 2001US-0314248P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Barnes G, Meyer J;
XX
WPI; 2002-627522/67.
XX
PT New isolated nucleic acid molecule with an allelic variant of a
PT polymorphic region of an 5-LO gene, useful for diagnosing and/or
PT prognosticating disorders associated with an aberrant inflammatory
PT response such as asthma.
XX
PS Claim 1; Page 234; 290pp; English.
XX
CC The invention relates to an isolated human nucleic acid molecule
CC comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
CC (5-LO) gene, where the allelic variant comprises one or more nucleotide
CC selected from any of 3, 20 or 21 base pair sequences, given in the
CC specification, or their complement. The compositions and methods of the
CC present invention are useful for diagnosing and/or prognosing disorders
CC associated with an aberrant inflammatory response such as asthma,
CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory

CC disease, atopic and contact dermatitis. The nucleic acid molecules can
CC also be useful for identifying an individual amongst other individuals
CC from the same species for use in forensic medicine and paternity testing.
CC This polynucleotide sequence represents DNA relating to the human 5-
CC lipoxigenase (5-LO) gene of the invention
XX
SQ Sequence 21 BP; 5 A; 7 C; 6 G; 3 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 6; Length 21;
Best Local Similarity 95.2%; Pred. No. 2; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 1;

Qy 1 AGGAGCGCGCAAAACCTTCTC 21
Db 1 AGGAGCGCGCAAAACCTTCTC 21

RESULT 3
AAT88433
ID AAT88433 standard; DNA; 2177 BP.
XX
AC AAT88433;
XX
DT 14-MAY-1998 (first entry)
XX
DE Human 5-lipoxygenase gene polymorphism 1669 to 1680 deletion.
XX
KW Inflammatory disease; polymorphism; 5-lipoxygenase; asthma;
KW ulcerative colitis; bronchitis; sinusitis; psoriasis; rhinitis;
KW arthritis; diagnosis; treatment; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
FN WO9742347-A2.
XX
PD 13-NOV-1997.
XX
PF 29-APR-1997; 97WO-US007137.
XX
PR 06-MAY-1996; 96US-0016890P.
XX
PR 25-APR-1997; 97US-00846020.
XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
XX
PI Drazen JM, In K, Asano K, Beier D, Grobholz J;
XX
WPI; 1997-558997/51.
XX
PT Classifying patients with inflammatory disease, specifically asthma -
PT according to polymorphisms in 5-lipoxygenase gene regulatory region, e.g.
PT to identify candidates for lipoxygenase inhibitor treatment.
XX
PS Claim 13; Page; 56pp; English.
XX
CC The present sequence was used in the development of a novel method for
CC classifying patients suffering from an inflammatory disease. The method
CC comprises identifying in DNA from at least 1 patient a sequence
CC polymorphism, as compared with the normal 5-lipoxygenase (5-LOX) gene
CC (AAT88431), in a 5-LOX regulatory gene sequence. The method can be
CC applied to subjects with asthma, ulcerative colitis, bronchitis,
CC sinusitis, psoriasis, allergic and non-allergic rhinitis, lupus or
CC rheumatoid arthritis. Specifically it can be used to diagnose asthma or
CC susceptibility to disease, identify treatments suitable for individual
CC patients or assess the likely success of treatment. N.B. Sequence not
CC given in the specification, but constructed using the wild type human 5-
CC lipoxigenase gene sequence given on pages 40 to 41
XX
SQ Sequence 2177 BP; 541 A; 597 C; 654 G; 385 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 2; Length 2177;
Best Local Similarity 95.2%; Pred. No. 3.7; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 1;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 990 AGGAGCGCGCGAAACCTTCTC 1010

RESULT 4
AAT88432
ID AAT88432 standard; DNA; 2183 BP.

AC AAT88432;

XX 14-MAY-1998 (first entry)

XX Human 5-lipoxygenase gene polymorphism 1669 to 1674 deletion.

XX Inflammatory disease; polymorphism; 5-lipoxygenase; asthma;
KW ulcerative colitis; bronchitis; sinusitis; psoriasis; rhinitis;
KW arthritis; diagnosis; treatment; ds.

XX Homo sapiens.

OS Synthetic

XX WO9742347-A2.

XX 13-NOV-1997.

XX 29-APR-1997; 97WO-US0071137.

XX 06-MAY-1996; 96US-0016890P.

XX 25-APR-1997; 97US-00846020.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL.

XX Drazen JM, In K, Asano K, Beier D, Grobholz J;

XX WPI; 1997-558997/51.

XX Classifying patients with inflammatory disease, specifically asthma -
PT according to polymorphisms in 5-lipoxygenase gene regulatory region, e.g.
PT to identify candidates for lipoxygenase inhibitor treatment.

XX Claim 12; Page; 56pp; English.

XX The present sequence was used in the development of a novel method for
CC classifying patients suffering from an inflammatory disease. The method
CC comprises identifying in DNA from at least 1 patient a sequence
CC polymorphism, as compared with the normal 5-lipoxygenase (5-LOX) gene
CC (AAT88431), in a 5-LOX regulatory gene sequence. The method can be
CC applied to subjects with asthma, ulcerative colitis, bronchitis,
CC sinusitis, psoriasis, allergic and non-allergic rhinitis, lupus or
CC rheumatoid arthritis. Specifically it can be used to diagnose asthma or
CC susceptibility to disease. Specifically it can be used to diagnose asthma or
CC patients or assess the likely success of treatment. N.B. Sequence not
CC given in the specification, but constructed using the wild type human 5-
CC lipoxygenase gene sequence given on pages 40 to 41

XX Sequence 2183 BP; 541 A; 598 C; 659 G; 385 T; 0 U; 0 Other;

XX Query Match 92.4%; Score 19.4; DB 2; Length 2183;
XX Best Local Similarity 95.2%; Pred. No. 3.7;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21

Db 990 AGGAGCGCGCGAAACCTTCTC 1010

RESULT 5

AAT88431

ID AAT88431 standard; DNA; 2189 BP.

XX AAT88431;

XX 14-MAY-1998 (first entry)
XX Human 5-lipoxygenase gene.

XX Inflammatory disease; polymorphism; 5-lipoxygenase; asthma;
KW ulcerative colitis; bronchitis; sinusitis; psoriasis; rhinitis;
KW arthritis; diagnosis; treatment; ds.

XX Homo sapiens.

XX WO9742347-A2.

XX 13-NOV-1997.

XX 29-APR-1997; 97WO-US0071137.

XX 06-MAY-1996; 96US-0016890P.

XX 25-APR-1997; 97US-00846020.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL.

XX Drazen JM, In K, Asano K, Beier D, Grobholz J;

XX WPI; 1997-558997/51.

XX Classifying patients with inflammatory disease, specifically asthma
PT according to polymorphisms in 5-lipoxygenase gene regulatory region, e.g.
PT to identify candidates for lipoxygenase inhibitor treatment.

XX Claim 11; Page 40-41; 56pp; English.

XX The present sequence was used in the development of a novel method for
CC classifying patients suffering from an inflammatory disease. The method
CC comprises identifying in DNA from at least 1 patient a sequence
CC polymorphism, as compared with the normal 5-lipoxygenase (5-LOX) gene
CC (AAT88431), in a 5-LOX regulatory gene sequence. The method can be
CC applied to subjects with asthma, ulcerative colitis, bronchitis,
CC sinusitis, psoriasis, allergic and non-allergic rhinitis, lupus or
CC rheumatoid arthritis. Specifically it can be used to diagnose asthma or
CC susceptibility to disease. Specifically it can be used to diagnose asthma or
CC patients or assess the likely success of treatment

XX Sequence 2189 BP; 541 A; 599 C; 664 G; 385 T; 0 U; 0 Other;

XX Query Match 92.4%; Score 19.4; DB 2; Length 2189;
XX Best Local Similarity 95.2%; Pred. No. 3.7;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21

Db 990 AGGAGCGCGCGAAACCTTCTC 1010

RESULT 6

AAT11113

ID AAT11113 standard; DNA; 2189 BP.

XX AAT11113;

XX 05-DEC-2002 (first entry)

XX Human 5-lipoxygenase gene related DNA sequence SEQ ID No 1.

XX Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
KW ds.

XX Homo sapiens.

XX

found teaching
specific one

```

PN WO200262825-A2.
XX
XX
PD 15-AUG-2002.
XX
XX
PF 07-FEB-2002; 2002WO-US003546.
XX
XX
PR 08-FEB-2001; 2001US-0267515P.
PR 21-AUG-2001; 2001US-0314248P.
XX
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
XX
PI Barnes G, Meyer J;
XX
XX
DR WPI; 2002-627522/67.
XX
XX
PT New isolated nucleic acid molecule with an allelic variant of a
PT polymorphic region of an 5-LO gene, useful for diagnosing and/or
PT prognosticating disorders associated with an aberrant inflammatory
PT response such as asthma.
XX
PS Claim 6; Fig 1; 290pp; English.
XX
CC The invention relates to an isolated human nucleic acid molecule
CC comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
CC (5-LO) gene, where the allelic variant comprises one or more nucleotide
CC selected from any of 3, 20 or 21 base pair sequences, given in the
CC specification, or their complement. The compositions and methods of the
CC present invention are useful for diagnosing and/or prognosing disorders
CC associated with an aberrant inflammatory response such as asthma,
CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
CC disease, atopic and contact dermatitis. The nucleic acid molecules can
CC also be useful for identifying an individual amongst other individuals
CC from the same species for use in forensic medicine and paternity testing.
CC This polynucleotide sequence represents DNA relating to the human 5-
CC lipoxygenase (5-LO) gene of the invention
XX
SQ Sequence 2189 BP; 541 A; 599 C; 664 G; 385 T; 0 U; 0 Other;
Query Match 92.4%; Score 19.4; DB 6; Length 2189;
Best Local Similarity 95.2%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 990 AGGAGCGCGCAAAACCTTCTC 1010
RESULT 7
AAD24657
ID AAD24657 standard; DNA; 2189 BP.
XX
XX
AC AAD24657;
XX
XX
DT 12-MAR-2002 (first entry)
XX
DE Human 5-lipoxygenase (ALOX5) gene partial coding sequence.
XX
XX Human; 5-lipoxygenase; ALOX5; respiratory disease; therapy; asthma;
XX leukotriene receptor antagonist; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH repeat_region /*tag= a
FT /*rpt_type= TANDEM
FT repeat_unit /*tag= b
FT /*note= "S1 binding motif (GGCGG)"
FT misc_signal 1845..1847
FT /*tag= c

```

```

FT WO200179560-A2.
XX
XX
PN 25-OCT-2001.
XX
XX
PF 17-APR-2001; 2001WO-US012534.
XX
XX
PR 17-APR-2000; 2000US-0197913P.
PR 29-SEP-2000; 2000US-0236608P.
XX
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
XX
PI Anderson WH, Edwards LD, Emmett AH, Pillai S, Sprankel CS;
XX
XX
DR WPI; 2002-066375/09.
XX
XX
PT Screening a subject suffering from a respiratory disease that can be
PT treated with a leukotriene receptor antagonist, by determining the
PT genotype of the DNA at a polymorphic allele site in the 5-lipoxygenase
PT gene and LTC4 synthase gene.
XX
XX
PS Example 1; Page 12-13; 67pp; English.
XX
CC The invention relates to a method of screening a subject suffering from a
CC respiratory disease which can be treated with a leukotriene receptor
CC antagonist. The method comprises determining the genotype of the DNA at a
CC polymorphic allele site in the 5-lipoxygenase gene and a polymorphic
CC allelic site in the LTC4 synthase gene, where different genotypes at
CC these sites are associated with different incidences of a phenotypic
CC response to the treatment. The method is useful for screening a subject
CC suffering from a respiratory disease, which can be treated with a
CC leukotriene receptor antagonist, particularly asthma, as an aid in
CC predicting the subject's response to the treatment. The present sequence
CC is human 5-lipoxygenase (ALOX5) gene partial coding sequence
XX
SQ Sequence 2189 BP; 541 A; 599 C; 664 G; 385 T; 0 U; 0 Other;
Query Match 92.4%; Score 19.4; DB 6; Length 2189;
Best Local Similarity 95.2%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 990 AGGAGCGCGCAAAACCTTCTC 1010
RESULT 8
AAT88434
ID AAT88434 standard; DNA; 2195 BP.
XX
XX
AC AAT88434;
XX
XX
DT 14-MAY-1998 (first entry)
XX
DE Human 5-lipoxygenase gene polymorphism 1688 (GGCGG) 6 bp addition.
XX
XX Inflammatory disease; polymorphism; 5-lipoxygenase; asthma;
XX ulcerative colitis; bronchitis; sinusitis; psoriasis; rhinitis;
XX arthritis; diagnosis; treatment; ds.
XX
OS Homo sapiens.
XX
XX Synthetic.
XX
XX PN WO9742347-A2.
XX
XX
PD 13-NOV-1997.
XX
XX
PF 29-APR-1997; 97WO-US007137.
XX
XX
PR 06-MAY-1996; 96US-0016890P.
PR 25-APR-1997; 97US-00846020.
XX
XX

```


comprising an allelic variant of a polymorphic region of a 5-lipoxygenase (5-LO) gene, where the allelic variant comprises one or more nucleotide selected from any of 3, 20 or 21 base pair sequences, given in the specification, or their complement. The compositions and methods of the present invention are useful for diagnosing and/or prognosing disorders associated with an aberrant inflammatory response such as asthma, bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis, rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory disease, atopic and contact dermatitis. The nucleic acid molecules can also be useful for identifying an individual amongst other individuals from the same species for use in forensic medicine and paternity testing. This polynucleotide sequence represents DNA relating to the human 5-lipoxygenase (5-LO) gene of the invention

Sequence 168273 BP; 46834 A; 36467 C; 36966 G; 46498 T; 0 U; 1508 Other;

Query Match 92.4%; Score 19.4; DB 6; Length 168273;
 Best Local Similarity 95.2%; Pred. No. 6.6;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
 |||||
 Db 167457 AGGAGCGCGCAAAACCTTCTC 167477

RESULT 11
 ABZ39218/c
 ID ABZ39218 standard; DNA; 975 BP.

AC ABZ39218;
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE N. gonorrhoeae nucleotide sequence SEQ ID 3025.

XX Antibacterial; infection; vaccine; gene therapy; gene; ds.

OS Neisseria gonorrhoeae.

XX WO200279243-A2.

PD 10-OCT-2002.

PF 12-FEB-2002; 2002WO-IB002069.

XX 12-FEB-2001; 2001GB-00003424.

PA (CHIR-) CHIRON SPA.

PI Fontana MR, Pizza M, Masignani V, Monaci E;

XX WPI; 2003-058415/05.

DR P-PSDB; ABP78248.

PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection.

PS Disclosure; Page 409; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae. Also disclosed are the nucleic acid molecules encoding the proteins and antibodies that specifically bind to the proteins. The composition comprising the protein, nucleic acid or antibody is useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection, this may be in the form of a vaccine or gene therapy. Sequences given in records ABZ37706-ABZ42016 represent nucleic acid molecules of the invention

XX Sequence 975 BP; 223 A; 226 C; 292 G; 234 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 7; Length 975;
 Best Local Similarity 90.0%; Pred. No. 80;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AGGAGCGCGCAAAACCTTCT 20
 |||||
 Db 516 AGGAGCGCGCAAAACCTTCT 497

RESULT 12

ACA40953/c

ID ACA40953 standard; DNA; 975 BP.

XX ACA40953;

DT 27-OCT-2003 (revised)

DT 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #22610.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.

OS Neisseria gonorrhoeae.

XX WO200277183-A2.

PD 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU37083.

PT New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 28823; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway; (8) required for proliferation, or that inhibits cellular proliferation; (9) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)

XX Sequence 975 BP; 223 A; 226 C; 292 G; 234 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 7; Length 975;
Best Local Similarity 90.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGGCGCAAAACCTTCT 20
Db 516 AGGAGCGGCGCAAAACCTCT 497

RESULT 13

AAA81505/c
ID AAA81505 standard; DNA; 28874 BP.

XX AAA81505;

DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm_52 SEQ ID NO:52.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; Gs.

OS Neisseria meningitidis.

PN WO2000022430-A2.

XX 20-APR-2000.

PF 08-OCT-1999; 99WO-US023573.

XX 09-OCT-1998; 98US-0103794P.

PR 30-APR-1999; 99US-0132068P.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarcellini M, Scarlato V;
PI Rappuoli R, Pizza M;

DR WPI; 2000-318079/27.

XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used
PT in the diagnosis and treatment of *N. meningitidis* infection and other
PT Neisserial infections, for example, *N. gonorrhoea*.

XX Claim 7; Page 1351-1359; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic proteins
CC from *Neisseria* genomic sequences. AAA81453 to AAA8414 represent
CC specifically claimed *Neisseria meningitidis* genomic DNA sequences;
CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA
CC sequences and their corresponding proteins; AAA81254 to AAA81259 and
CC AAA81304 to AAA81321 represent PCR primers used in the isolation of
CC *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent
CC *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all
CC used in the exemplification of the present invention. The nucleic acid
CC sequences, protein sequences, and antibodies against them, can be used in
CC the manufacture of a composition. The composition can be used as a
CC medicament (or in the manufacture of a medicament) for treating,
CC preventing or diagnosing infection due to *Neisserial* bacteria. For
CC example, some of the identified proteins could be components of vaccines

CC against *Meningococcus* B; against all serotypes; and/or against all
CC pathogenic *Neisseriae*. Identification of sequences from the bacterium
CC will also facilitate production of biological probes, particularly
CC organism-specific probes. Attempts to make efficacious *Meningococcus* B
CC vaccines have failed mainly due to antigen tolerance. Multivalent
CC vaccines have also been tried but none have successfully overcome
CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
CC variable regions

XX Sequence 28874 BP; 6553 A; 7289 C; 8222 G; 6810 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 3; Length 28874;

Best Local Similarity 90.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGGCGCAAAACCTTCT 20

Db 11903 AGGAGCGGCGCAAAACCTCT 11884

RESULT 14

AAA81490_07/c

Continuation (8 of 15) of AAA81490 from base 700001 (*N. meningitidis* B full length gene
WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aaa81490

WP Fragment Name Begin End

WP AAA81490_00 1 110000
WP AAA81490_01 100001 210000
WP AAA81490_02 200001 310000
WP AAA81490_03 300001 410000
WP AAA81490_04 400001 510000
WP AAA81490_05 500001 610000
WP AAA81490_06 600001 710000
WP AAA81490_07 700001 810000
WP AAA81490_08 800001 910000
WP AAA81490_09 900001 1010000
WP AAA81490_10 1000001 1110000
WP AAA81490_11 1100001 1210000
WP AAA81490_12 1200001 1310000
WP AAA81490_13 1300001 1410000
WP AAA81490_14 1400001 1437668

Query Match 80.0%; Score 16.8; DB 3; Length 110000;

Best Local Similarity 90.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGGCGCAAAACCTTCT 20

Db 98755 AGGAGCGGCGCAAAACCTCT 98736

RESULT 15

AAF21608/c

ID AAF21608 standard; DNA; 349980 BP.

XX AAF21608;

DT 13-MAR-2001 (first entry)

XX *Neisseria meningitidis* B nucleotide sequence SEQ ID NO:109.

XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.

XX *Neisseria meningitidis*.

XX WO200006791-A1.

XX 09-NOV-2000.

XX

```

PF 08-MAR-2000; 2000WO-US005928.
XX
PR 30-APR-1999; 99US-0132068P.
PR 08-OCT-1999; 99WO-US023573.
PR 28-FEB-2000; 2000GB-00004695.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections.
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of Neisseria
CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
CC represent fragments of the NMB genomic sequence, as the sequence was too
CC long to go in a record on its own it was split into 8 sequences which
CC overlap each other at the beginning and end of each sequence by 49980 bp
CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
CC primers which are used in the exemplification of the present invention.
CC The NMB genome and fragments from it have antibacterial activity, and can
CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
CC and/or antibodies which binds to the proteins can be used in compositions
CC for treating or preventing infection due to Neisserial bacteria or as a
CC diagnostic reagent for detecting the presence of Neisserial bacteria or
CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
CC computer storage medium or computer databases can be used in a search to
CC identify open reading frames (ORFs) or coding sequences within the NMB
CC genome. The DNA sequences provide further opportunities to find antigenic
CC or immunogenic proteins which are more effective in vaccines than the
CC outer membrane proteins currently used
XX
SQ Sequence 349980 BP; 82523 A; 82940 C; 96712 G; 87805 T; 0 U; 0 Other;
Query Match 80.0%; Score 16.8; DB 3; Length 349980;
Best Local Similarity 90.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGGAGCGCGCAAAACCTTCT 20
DB 198756 AGGAGCGCGCAAAACCATCT 198737

```

Search completed: May 7, 2004, 10:10:26
Job time : 153.048 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 10:10:37 ; Search time 35,258 Seconds
(without alignments)
330.836 Million cell updates/sec

Title: US-10-071-411A-4

Perfect score: 21

Sequence: 1 agagcgcgcaaacctctc 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	2189	3	US-08-846-020A-1
2	19.4	92.4	2189	4	US-09-617-871-1
3	16.2	77.1	2051	1	US-08-343-785-7
4	16.2	77.1	2051	2	US-08-462-221-7
5	16.2	77.1	2051	3	US-08-946-458-7
6	16.2	77.1	2111	1	US-08-343-785-1
7	16.2	77.1	2111	2	US-08-462-221-1
8	16.2	77.1	2111	3	US-08-946-458-1
9	15.8	75.2	2988	4	US-09-657-931A-2
10	15.4	73.3	1830121	4	US-09-557-884-3
11	15.4	73.3	1830121	4	US-09-643-990A-1
12	15.2	72.4	417	4	US-09-252-991A-2715
13	15.2	72.4	881	1	US-08-631-200-12
14	15.2	72.4	881	1	US-08-829-555-12
15	15.2	72.4	881	2	US-08-922-267A-12
16	15.2	72.4	881	2	US-08-936-707A-12
17	15.2	72.4	881	2	US-08-936-706A-12
18	15.2	72.4	881	3	US-09-248-203-12
19	15.2	72.4	881	3	US-09-406-071-12
20	15.2	72.4	881	4	US-09-814-986-12
21	15.2	72.4	1338	3	US-09-027-064-3
22	15.2	72.4	1338	3	US-09-271-815-3
23	15.2	72.4	1899	4	US-09-328-352-659
24	15.2	72.4	2394	3	US-09-027-064-1
25	15.2	72.4	2394	3	US-09-271-815-1
26	15.2	72.4	2427	4	US-09-252-991A-2899
27	15.2	72.4	2556	4	US-09-252-991A-3074

28	15.2	72.4	2671	2	US-08-408-519-1	Sequence 1, Appli
29	15.2	72.4	2671	5	PCT-US95-03552-1	Sequence 1, Appli
30	15.2	72.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
31	15.2	72.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
32	14.8	70.5	293	4	US-09-313-294A-3012	Sequence 3012, Ap
33	14.8	70.5	1173	4	US-09-489-039A-1206	Sequence 1206, Ap
34	14.8	70.5	2762	1	US-08-198-446B-12	Sequence 12, Appl
35	14.8	70.5	2762	2	US-08-870-693-12	Sequence 12, Appl
36	14.8	70.5	4899	6	5268270-1	Patent No. 5268270
37	14.8	70.5	5956	4	US-09-112-580-12	Sequence 12, Appl
38	14.6	69.5	470	4	US-09-621-976-1321	Sequence 1321, Ap
39	14.6	69.5	851	3	US-09-126-646-3	Sequence 3, Appli
40	14.6	69.5	851	4	US-09-421-491-3	Sequence 3, Appli
41	14.6	69.5	999	4	US-08-436-630-31	Sequence 31, Appl
42	14.6	69.5	1215	4	US-09-543-681A-176	Sequence 176, App
43	14.6	69.5	1487	1	US-07-865-662F-11	Sequence 11, Appl
44	14.6	69.5	1487	3	US-08-374-219B-11	Sequence 11, Appl
45	14.6	69.5	1498	2	US-08-818-034-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-846-020A-1
; Sequence 1, Application US/08846020A
; Patent No. 6090547
; GENERAL INFORMATION:
; APPLICANT: Drzen M.D., Jeffrey M.
; APPLICANT: In M.D., Kwang-Ho
; APPLICANT: Asano M.D., Koichiro
; APPLICANT: Beier, David
; APPLICANT: Grobholz, James
; TITLE OF INVENTION: 5-Lipoxygenase Gene Sequence
; TITLE OF INVENTION: Polymorphisms and Their Use in Classifying Patients
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHOATE, HALL & STEWART
; STREET: 53 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2891
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,020A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Jarrell Ph.D., Brenda H.
; REGISTRATION NUMBER: 39,223
; REFERENCE/DOCKET NUMBER: 0092662-0012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-5000
; TELEFAX: (617) 248 4000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2189 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: human 5-lipoxygenase gene (GenBank M 38191)
US-08-846-020A-1

Query Match 92.4%; Score 19.4; DB 3; Length 2189;
Best Local Similarity 95.2%; Pred. No. 0.45;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```
Qy 1 AGGAGCGCGCAAAACCTTCTC 21
Db 990 AGGAGCGCGCAAAACCTTCTC 1010

RESULT 2
US-09-617-871-1
; Sequence 1, Application US/09617871
; Patent No. 6355434
; GENERAL INFORMATION:
; APPLICANT: Drzen M.D., Jeffrey M.
; APPLICANT: In M.D., Kwang-Ho
; APPLICANT: Asano M.D., Koichiro
; APPLICANT: Beier, David
; APPLICANT: Grobholz, James
; TITLE OF INVENTION: 5-Lipoxygenase Gene Sequence
; TITLE OF INVENTION: Polymorphisms and Their Use in Classifying Patients
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHOATE, HALL & STEWART
; STREET: 53 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2891
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/617,871
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/846,020
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jarrell Ph.D., Brenda H.
; REGISTRATION NUMBER: 39,223
; REFERENCE/DOCKET NUMBER: 0092662-0012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-5000
; TELEFAX: (617) 248 4000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2189 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: human 5-lipoxygenase gene (GenBank M 38191)
; US-09-617-871-1

Query Match 92.4%; Score 19.4; DB 4; Length 2189;
Best Local Similarity 95.2%; Pred. No. 0.45;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGAGCGCGCAAAACCTTCTC 21
Db 990 AGGAGCGCGCAAAACCTTCTC 1010

RESULT 3
US-08-343-785-7
; Sequence 7, Application US/08343785
; Patent No. 5494810
; GENERAL INFORMATION:
; APPLICANT: Francis Barany et al
; TITLE OF INVENTION: Thermostable Ligase Mediated DNA
; TITLE OF INVENTION: Amplification System For The
```

```
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,785
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/971,095
; FILING DATE: NO. 5494810ember 2nd 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; REFERENCE/DOCKET NUMBER: CRF D-933/CIT-2047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-343-785-7

Query Match 77.1%; Score 16.2; DB 1; Length 2051;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AGGAGCGCGCAAAACCTTCTC 21
Db 1474 AAGAGCGCGCAAAACCTTCTC 1494

RESULT 4
US-08-462-221-7
; Sequence 7, Application US/08462221
; Patent No. 5830711
; GENERAL INFORMATION:
; APPLICANT: Barany, Francis
; APPLICANT: Zebala, John
; APPLICANT: Nickerson, Deborah
; APPLICANT: Kaiser, Jr., Robert J.
; APPLICANT: Hood, Leroy
; TITLE OF INVENTION: A THERMOSTABLE LIGASE MEDIATED DNA
; TITLE OF INVENTION: AMPLIFICATION SYSTEM FOR THE DETECTION OF GENETIC
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,221
; FILING DATE: 05-JUN-1995
```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/343,785
FILING DATE: 22-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,095
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,447
FILING DATE: 03-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/267
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1304
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2051 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-462-221-7

Query Match 77.1%; Score 16.2; DB 2; Length 2051;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 1474 AAGAGCGCCCAAAACCTCTC 1494

RESULT 5
US-08-946-458-7
Sequence 7, Application US/08946458
Patent No. 6054564
GENERAL INFORMATION:
APPLICANT: Barany, Francis
APPLICANT: Zebala, John
APPLICANT: Nickerson, Deborah
APPLICANT: Kaiser, Jr., Robert J.
APPLICANT: Hood, Leroy
TITLE OF INVENTION: A THERMOSTABLE LIGASE MEDIATED DNA
AMPLIFICATION SYSTEM FOR THE DETECTION OF GENETIC DISEASES
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon Peabody LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,458
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/462,221
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/343,785
FILING DATE: 22-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,095
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/518,447
FILING DATE: 03-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/267
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1304
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2051 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-946-458-7

Query Match 77.1%; Score 16.2; DB 3; Length 2051;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 1474 AAGAGCGCCCAAAACCTCTC 1494

RESULT 6
US-08-343-785-1
Sequence 1, Application US/08343785
Patent No. 5494810
GENERAL INFORMATION:
APPLICANT: Francis Barany et al
TITLE OF INVENTION: Thermostable Ligase Mediated DNA
AMPLIFICATION SYSTEM FOR THE
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Yahwak & Associates
STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,785
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/971,095
FILING DATE: No. 5494810ember 2nd 1992
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-933/CIT-2047
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2111 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-343-785-1

Query Match 77.1%; Score 16.2; DB 1; Length 2111;
Best Local Similarity 85.7%; Pred. No. 23;

; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2988
; TYPE: DNA
; ORGANISM: Manduca sexta
US-09-657-931A-2

Query Match 75.2%; Score 15.8; DB 4; Length 2988;
Best Local Similarity 89.5%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGAGCGCGCAAAACCTTC 19
||| ||||| ||||| |||||
DB 1171 AGAAGCGCGCAAAACCTTC 1153

RESULT 10

US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 73.3%; Score 15.4; DB 4; Length 1830121;
Best Local Similarity 94.1%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AGCGCGCAAAACCTTCT 20
||||| ||||| ||||| |||||
DB 1762409 AGCGCGCAAAACCTTCT 1762393

RESULT 11

US-09-643-990A-1/c

; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:

; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter

; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 73.3%; Score 15.4; DB 4; Length 1830121;
Best Local Similarity 94.1%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AGCGCGCAAAACCTTCT 20
||||| ||||| ||||| |||||
DB 1762409 AGCGCGCAAAACCTTCT 1762393

RESULT 12

US-09-252-991A-2715
; Sequence 2715, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27

```
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2715
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2715

Query Match          72.4%; Score 15.2; DB 4; Length 417;
Best Local Similarity 85.0%; Pred. No. 59;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGAGCGCGCAAAACCTTCT 21
Db 151 GAAGCGCGCAAGCCTTCT 170

RESULT 13
US-08-631-200-12/c
; Sequence 12, Application US/08631200
; Patent No. 5646040
; GENERAL INFORMATION:
; APPLICANT: Kleyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,553
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/631,200
; FILING DATE: 12-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-829-553-12

Query Match          72.4%; Score 15.2; DB 1; Length 881;
Best Local Similarity 85.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCT 20
Db 502 AGGAGCGCGCAAAACATCT 483

RESULT 15
US-08-922-267A-12/c
; Sequence 12, Application US/08922267A
; Patent No. 5861239
; GENERAL INFORMATION:
; APPLICANT: Kleyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/922,267A
; FILING DATE: 2-SEP-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/829,553
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/631,200
; FILING DATE: 12-APR-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-922-267A-12

```

```

Query Match          72.4%; Score 15.2; DB 2; Length 881;
Best Local Similarity 85.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 AGGAGCGCGCAAAACCTTCT 20
        ||||| ||||| |||||
Db      502 AGGAGCCTGCAAAACATTTCT 483

```

```

Search completed: May 7, 2004, 11:56:19
Job time : 45.2258 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 11:54:38 ; Search time 148.016 Seconds
(without alignments)
642.699 Million cell updates/sec

Title: US-10-071-411a-4

Perfect score: 21

Sequence: 1 aggagcgcgaacattctc 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	2189	15	US-10-240-305-5
2	16.8	80.0	975	13	US-10-282-122A-28823
3	16.8	80.0	1392	16	US-10-369-493-33373
4	16.2	77.1	1577	13	US-10-425-114-31900
5	16.2	77.1	2051	13	US-10-662-199-7
6	16.2	77.1	2111	13	US-10-662-199-1
7	16.2	77.1	1163020	16	US-10-398-221-10
8	16.2	77.1	3011208	16	US-10-398-221-2058
9	16	76.2	9899	15	US-10-311-455-2416
10	16	76.2	3673778	15	US-10-312-841-1
11	15.8	75.2	592	10	US-09-918-995-2700
12	15.8	75.2	1260	16	US-10-369-493-35063
13	15.8	75.2	1260	16	US-10-369-493-38465
14	15.8	75.2	1260	16	US-10-369-493-38619

C	15	15.8	75.2	1260	16	US-10-369-493-38956	Sequence 38956, A
	16	15.8	75.2	1597	13	US-10-424-598-3431	Sequence 3431, Ap
	17	15.8	75.2	1717	13	US-10-424-598-3435	Sequence 3435, Ap
	18	15.8	75.2	2427	13	US-10-424-598-3435	Sequence 3434, Ap
	19	15.8	75.2	116422	13	US-10-087-192-1195	Sequence 1195, Ap
	20	15.4	73.3	352	10	US-09-803-719-1281	Sequence 1281, Ap
	21	15.4	73.3	451	9	US-09-983-965-1333	Sequence 1333, Ap
C	22	15.4	73.3	777	16	US-10-369-493-44731	Sequence 44731, A
	23	15.4	73.3	1188	13	US-10-282-122A-24882	Sequence 24882, A
C	24	15.4	73.3	1830121	15	US-10-329-960-1	Sequence 1, Appli
C	25	15.4	73.3	1830121	16	US-10-329-960-1	Sequence 1, Appli
C	26	15.2	72.4	60	10	US-09-908-975-15883	Sequence 15883, A
	27	15.2	72.4	371	13	US-10-424-598-47189	Sequence 47189, A
	28	15.2	72.4	451	11	US-09-864-408A-371	Sequence 371, Appl
C	29	15.2	72.4	455	10	US-09-918-995-27605	Sequence 27605, A
	30	15.2	72.4	466	10	US-09-918-995-5452	Sequence 5452, Ap
C	31	15.2	72.4	500	15	US-10-029-386-11377	Sequence 11377, A
C	32	15.2	72.4	500	15	US-10-029-386-11419	Sequence 11419, A
	33	15.2	72.4	508	9	US-09-974-300-7404	Sequence 7404, Ap
C	34	15.2	72.4	519	13	US-10-027-632-289910	Sequence 289910, A
C	35	15.2	72.4	519	16	US-10-027-632-289910	Sequence 289910, A
C	36	15.2	72.4	544	15	US-10-029-386-8915	Sequence 8915, Ap
C	37	15.2	72.4	563	13	US-10-027-632-280313	Sequence 280313, A
	38	15.2	72.4	563	16	US-10-027-632-280313	Sequence 280313, A
C	39	15.2	72.4	574	10	US-09-986-480-33	Sequence 33, Appl
	40	15.2	72.4	575	10	US-09-814-353-14752	Sequence 14752, A
C	41	15.2	72.4	578	13	US-10-027-632-224659	Sequence 224659, A
C	42	15.2	72.4	578	13	US-10-027-632-224659	Sequence 224659, A
C	43	15.2	72.4	578	16	US-10-027-632-224660	Sequence 224660, A
C	44	15.2	72.4	578	16	US-10-027-632-224660	Sequence 224660, A
C	45	15.2	72.4	588	16	US-10-369-493-34044	Sequence 34044, A

ALIGNMENTS

RESULT 1
US-10-240-305-5
; Sequence 5, Application US/10240305
; Publication No. US20030162193A1
; GENERAL INFORMATION:
; APPLICANT: GLAXO GROUP LIMITED
; APPLICANT: ANDERSON, WAYNE H.
; APPLICANT: EDWARDS, Lisa D.
; APPLICANT: EMMETT, Amanda H.
; APPLICANT: PILLAI, Sreekumar
; APPLICANT: SPRANKEL, Catherine S.
; TITLE OF INVENTION: Medicine Response Assay in Respiratory Disease
; FILE REFERENCE: PU3958 & PU4254
; CURRENT APPLICATION NUMBER: US/10/240,305
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-240-305-5

Query Match 92.4%; Score 19.4; DB 15; Length 2189;
Best Local Similarity 95.2%; Pred. No. 3.2;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAACTTCTC 21
|||||
Db 990 AGGAGCGCGCAAACTTCTC 1010

RESULT 2

US-10-282-122A-28823/c
; Sequence 28823, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

102


```
; PRIOR APPLICATION NUMBER: 09/480,515
; PRIOR FILING DATE: 2000-01-10
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Thermus aquaticus ligase
US-10-662-199-7

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 13; Length 2051;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 1474 AAGAGCGCGCAAAACCTTCTC 1494

RESULT 6
US-10-662-199-1
; Sequence 1, Application US/10662199
; Publication No. US20040048308A1
; GENERAL INFORMATION:
; APPLICANT: Barany, Francis
; APPLICANT: Zebala, John
; APPLICANT: Nickerson, Deborah
; APPLICANT: Kaiser Jr., Robert J.
; APPLICANT: Hood, Leroy
; TITLE OF INVENTION: A THERMOSTABLE LIGASE MEDIATED DNA AMPLIFICATION SYSTEM
; FILE REFERENCE: 19603/3641
; CURRENT APPLICATION NUMBER: US/10/662,199
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 08/946,458
; PRIOR FILING DATE: 1997-10-07
; PRIOR APPLICATION NUMBER: US 08/462,221
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: US 08/343,785
; PRIOR FILING DATE: 1994-11-22
; PRIOR APPLICATION NUMBER: US 07/971,095
; PRIOR FILING DATE: 1992-11-02
; PRIOR APPLICATION NUMBER: US 07/518,447
; PRIOR FILING DATE: 1990-05-03
; PRIOR APPLICATION NUMBER: 09/480,515
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2111
; TYPE: DNA
; ORGANISM: Thermus aquaticus ligase
US-10-662-199-1

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 13; Length 2111;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 1534 AAGAGCGCGCAAAACCTTCTC 1554

RESULT 7
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2001-07-02
```

```
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1163020
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 16; Length 1163020;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 776366 AGGAGCTCGGAAATCTCTC 776386

RESULT 8
US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
US-10-398-221-2058

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 16; Length 3011208;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
Db 866632 AGGAGCTCGGAAATCTCTC 866652

RESULT 9
US-10-311-455-2416/c
; Sequence 2416, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
```

;/ PRIOR APPLICATION NUMBER: DE 10032529.7
;/ PRIOR FILING DATE: 2000-06-30
;/ PRIOR APPLICATION NUMBER: DE 10043826.1
;/ PRIOR FILING DATE: 2000-09-01
;/ NUMBER OF SEQ ID NOS: 2424
;/ SEQ ID NO 2416
;/ LENGTH: 9899
;/ TYPE: DNA
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2416

Query Match 76.2%; Score 16; DB 15; Length 9899;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CGCGCAAAACCTTCTC 21
Db 7676 CGCGCAAAACCTTCTC 7661

RESULT 10
US-10-312-841-1/c
;/ Sequence 1, Application US/10312841
;/ Publication No. US20030186277A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Epigenomics AG
;/ TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;/ FILE REFERENCE: E01/1208/WO
;/ CURRENT APPLICATION NUMBER: US/10/312,841
;/ PRIOR FILING DATE: 2002-12-30
;/ NUMBER OF SEQ ID NOS: 2
;/ SEQ ID NO 1
;/ LENGTH: 3673778
;/ TYPE: DNA
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;/ FEATURE:
;/ NAME/KEY: unsure
;/ LOCATION: (3294164)
US-10-312-841-1

Query Match 76.2%; Score 16; DB 15; Length 3673778;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CGCGCAAAACCTTCTC 21
Db 2102473 CGCGCAAAACCTTCTC 2102458

RESULT 11
US-09-918-995-2700
;/ Sequence 2700, Application US/09918995
;/ Publication No. US20030073623A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Hyseq, Inc.
;/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;/ FILE REFERENCE: 20411-756
;/ CURRENT APPLICATION NUMBER: US/09/918,995
;/ CURRENT FILING DATE: 2001-07-30
;/ PRIOR APPLICATION NUMBER: US/09/235,076
;/ PRIOR FILING DATE: 1999-01-20
;/ NUMBER OF SEQ ID NOS: 38054
;/ SOFTWARE: PastSeq for Windows Version 3.0
;/ SEQ ID NO 2700
;/ LENGTH: 592
;/ TYPE: DNA
;/ ORGANISM: Homo sapiens
;/ FEATURE:

;/ NAME/KEY: misc feature
;/ LOCATION: (1) --(592)
;/ OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2700

Query Match 75.2%; Score 15.8; DB 10; Length 592;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTTC 19
Db 245 AGGCGCGCGCAAAACCTTTC 263

RESULT 12
US-10-369-493-35063/c
;/ Sequence 35063, Application US/10369493
;/ Publication No. US20030233675A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Cao, Yongwei
;/ APPLICANT: Hinkle, Gregory J.
;/ APPLICANT: Slater, Steven C.
;/ APPLICANT: Goldman, Barry S.
;/ APPLICANT: Chen, Xianfeng
;/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;/ FILE REFERENCE: 38-10(52052)B
;/ CURRENT APPLICATION NUMBER: US/10/369,493
;/ CURRENT FILING DATE: 2003-02-28
;/ PRIOR APPLICATION NUMBER: US 60/360,039
;/ PRIOR FILING DATE: 2002-02-21
;/ NUMBER OF SEQ ID NOS: 47374
;/ SEQ ID NO 35063
;/ LENGTH: 1260
;/ TYPE: DNA
;/ ORGANISM: Agrobacterium tumefaciens
US-10-369-493-35063

Query Match 75.2%; Score 15.8; DB 16; Length 1260;
Best Local Similarity 89.5%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTTC 19
Db 613 AGGCGAGCGCAAAACCTTTC 595

RESULT 13
US-10-369-493-38465/c
;/ Sequence 38465, Application US/10369493
;/ Publication No. US20030233675A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Cao, Yongwei
;/ APPLICANT: Hinkle, Gregory J.
;/ APPLICANT: Slater, Steven C.
;/ APPLICANT: Goldman, Barry S.
;/ APPLICANT: Chen, Xianfeng
;/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;/ FILE REFERENCE: 38-10(52052)B
;/ CURRENT APPLICATION NUMBER: US/10/369,493
;/ CURRENT FILING DATE: 2003-02-28
;/ PRIOR APPLICATION NUMBER: US 60/360,039
;/ PRIOR FILING DATE: 2002-02-21
;/ NUMBER OF SEQ ID NOS: 47374
;/ SEQ ID NO 38465
;/ LENGTH: 1260
;/ TYPE: DNA
;/ ORGANISM: Agrobacterium tumefaciens
US-10-369-493-38465

Query Match 75.2%; Score 15.8; DB 16; Length 1260;
Best Local Similarity 89.5%; Pred. No. 2.5e+02;

Search completed: May 7, 2004, 13:42:22
Job time : 165.016 secs

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTC 19
||| |||||
Db 613 AGCGAGCGCAAAACCTTC 595

RESULT 14

US-10-369-493-38619/c
; Sequence 38619, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 38619
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-38619

Query Match 75.2%; Score 15.8; DB 16; Length 1260;
Best Local Similarity 89.5%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTC 19
||| |||||
Db 613 AGCGAGCGCAAAACCTTC 595

RESULT 15

US-10-369-493-38956/c
; Sequence 38956, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 38956
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-38956

Query Match 75.2%; Score 15.8; DB 16; Length 1260;
Best Local Similarity 89.5%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTC 19
||| |||||
Db 613 AGCGAGCGCAAAACCTTC 595

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 09:56:42 ; Search time 1438.5 seconds
(without alignments)
435.944 Million cell updates/sec

Title: US-10-071-411A-4

Perfect score: 21

Sequence: 1 aggagcgcaaaccttctc 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba:*
2: em_esthum:*
3: em_estm:*
4: em_estm1:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.8	84.8	615	12	BJ174754
C 2	17.4	82.9	570	14	CA236358 SCACFL503
C 3	17.4	82.9	908	13	BUS69313 AGENCOURT
C 4	17	81.0	937	14	CD247125

	5	17	81.0	1121	13	BQ653696
C 6	16.8	80.0	325	14	CA782731	sat53d07.
C 7	16.8	80.0	379	29	CG905629	ZMVBb051
C 8	16.8	80.0	409	29	CG701116	ZMVBb011
C 9	16.8	80.0	427	29	CG878712	ZMVBb049
C 10	16.8	80.0	443	9	AA817025	LD22010.5
C 11	16.8	80.0	446	9	AA739254	vv51e01.r
C 12	16.8	80.0	567	29	CG692788	ZMVBb029
C 13	16.8	80.0	581	14	CA259725	SCRIPT302
C 14	16.8	80.0	584	28	BH486134	BOGDP74TR
C 15	16.8	80.0	608	28	CC101202	CSU-K34.1
C 16	16.8	80.0	618	9	AV658126	AV658126
C 17	16.8	80.0	619	29	FR0049585	Fugu rubr
C 18	16.8	80.0	721	28	BH663132	BOMNA74TF
C 19	16.8	80.0	733	28	BH476497	BOGH02TF
C 20	16.8	80.0	766	14	CF999763	AGENCOURT
C 21	16.8	80.0	826	28	BZ739181	AGENCOURT
C 22	16.8	80.0	831	13	BH853198	AGENCOURT
C 23	16.8	80.0	832	28	BH485282	BOGEG36TR
C 24	16.8	80.0	949	14	CF412665	CH3#081.B
C 25	16.8	80.0	1010	13	BH60191	AGENCOURT
C 26	16.4	78.1	597	10	AM645493	cms5e09.w
C 27	16.4	78.1	869	29	CG002656	ZUAE719TV
C 28	16.2	77.1	155	14	CD195849	MSI-0091T
C 29	16.2	77.1	186	14	CD196241	MSI-0092T
C 30	16.2	77.1	255	28	BH379820	AG-ND-141
C 31	16.2	77.1	317	12	BP516805	BP516805
C 32	16.2	77.1	351	9	AI613134	tw59q04.x
C 33	16.2	77.1	351	10	AM456666	UI-M-BH3-
C 34	16.2	77.1	355	9	AV748862	AV748862
C 35	16.2	77.1	374	12	BM867124	mgcs009xA
C 36	16.2	77.1	379	13	BY607480	BY607480
C 37	16.2	77.1	381	13	BY004145	BY004145
C 38	16.2	77.1	387	13	BY612588	BY612588
C 39	16.2	77.1	396	28	AQ125645	HS-2166.B
C 40	16.2	77.1	403	14	CA714829	wdk3c.pk0
C 41	16.2	77.1	406	13	BY633209	BY633209
C 42	16.2	77.1	409	14	CA716404	wdk3c.pk0
C 43	16.2	77.1	410	14	CB893403	CB893403
C 44	16.2	77.1	413	10	BE360864	DGI-67.A0
C 45	16.2	77.1	421	28	B89178	CIT-HSP-216

ALIGNMENTS

RESULT 1

BJ174754/c

LOCUS

DEFINITION

615 bp mRNA linear EST 16-OCT-2003

caulonemata and malformed buds Physcomitrella patens subsp. patens

CDNA clone pphb12p14 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Tadao Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'- GAGAGAGAGAGATCCACCTGGAGAGTTTTTTTTTTTNN-3' was used as a list 3' primer, and 5'-GTTCTCGAGTCATCGTGTCCAGACGATGCTGAGAACGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGCCAAACGCCAGCTCGAATTCGAGACCG). cDNA instext could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003). Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database of Physcomitrella EST clones is available at the PHYSCobase (http://moss.nibb.ac.jp/).	
FEATURES source	
1..615 Location/Qualifiers /organism="Physcomitrella patens subsp. patens" /mol_type="mRNA" /sub_species="patens" /db_xref="taxon:145481" /clone="ppb12p14" /tissue_type="mixture of chloronemata, caulonemata and malformed buds" /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"	
ORIGIN	
Query Match 84.8%; Score 17.8; DB 12; Length 615; Best Local Similarity 90.3%; Pred. No. 5.5e+02; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY 1 AGGAGCGCGCAAAACCTTCTC 21 	
Db 99 AGGATCGCGCAAAACCTTCTC 79 	
RESULT 2 CA236358 LOCUS DEFINITION CA236358 570 bp mRNA linear EST 25-SEP-2003 cDNA clone SCACFL5031A01 5', mRNA sequence.	
ACCESSION VERSION KEYWORDS SOURCE ORGANISM	
REFERENCE AUTHORS TITLE JOURNAL COMMENT	
FEATURES source	
1..908 Location/Qualifiers /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6616719" /lab_host="DH10B (T1 phage-resistant)" /clone_lib="NIH MGC 82" /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcccttcggcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGCCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length	
Query Match 82.9%; Score 17.4; DB 14; Length 570; Best Local Similarity 94.7%; Pred. No. 8.4e+02; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY 3 GAGCGCGCAAAACCTTCTC 21 	
Db 495 GAGCGCGCAACCTTCTC 513 	
RESULT 3 BU569313/c LOCUS DEFINITION BU569313 908 bp mRNA linear EST 16-SEP-2002 AGENCOURT_10366532 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6616719 5', mRNA sequence.	
ACCESSION VERSION KEYWORDS SOURCE ORGANISM	
REFERENCE AUTHORS TITLE JOURNAL COMMENT	
FEATURES source	
1..908 Location/Qualifiers /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6616719" /lab_host="DH10B (T1 phage-resistant)" /clone_lib="NIH MGC 82" /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcccttcggcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGCCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length	

TITLE	JOURNAL	AUTHORS
1. The Effect of the
2. The Effect of the
3. The Effect of the
4. The Effect of the
5. The Effect of the
6. The Effect of the
7. The Effect of the
8. The Effect of the
9. The Effect of the
10. The Effect of the
11. The Effect of the
12. The Effect of the
13. The Effect of the
14. The Effect of the
15. The Effect of the
16. The Effect of the
17. The Effect of the
18. The Effect of the
19. The Effect of the
20. The Effect of the
21. The Effect of the
22. The Effect of the
23. The Effect of the
24. The Effect of the
25. The Effect of the
26. The Effect of the
27. The Effect of the
28. The Effect of the
29. The Effect of the
30. The Effect of the
31. The Effect of the
32. The Effect of the
33. The Effect of the
34. The Effect of the
35. The Effect of the
36. The Effect of the
37. The Effect of the
38. The Effect of the
39. The Effect of the
40. The Effect of the
41. The Effect of the
42. The Effect of the
43. The Effect of the
44. The Effect of the
45. The Effect of the
46. The Effect of the
47. The Effect of the
48. The Effect of the
49. The Effect of the
50. The Effect of the
51. The Effect of the
52. The Effect of the
53. The Effect of the
54. The Effect of the
55. The Effect of the
56. The Effect of the
57. The Effect of the
58. The Effect of the
59. The Effect of the
60. The Effect of the
61. The Effect of the
62. The Effect of the
63. The Effect of the
64. The Effect of the
65. The Effect of the
66. The Effect of the
67. The Effect of the
68. The Effect of the
69. The Effect of the
70. The Effect of the
71. The Effect of the
72. The Effect of the
73. The Effect of the
74. The Effect of the
75. The Effect of the
76. The Effect of the
77. The Effect of the
78. The Effect of the
79. The Effect of the
80. The Effect of the
81. The Effect of the
82. The Effect of the
83. The Effect of the
84. The Effect of the
85. The Effect of the
86. The Effect of the
87. The Effect of the
88. The Effect of the
89. The Effect of the
90. The Effect of the
91. The Effect of the
92. The Effect of the
93. The Effect of the
94. The Effect of the
95. The Effect of the
96. The Effect of the
97. The Effect of the
98. The Effect of the
99. The Effect of the
100. The Effect of the

1 (bases 1 to 325)
Shoemaker, R., Keim, P., Vodka, L., Erpelding, J., Coryell, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estwatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com

Trace considered overall poor quality

Seq primer: -40KP from Gibco

High quality sequence stop: 1.

FEATURES

source
1. .325
Location/Qualifiers
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl056-5150"
/tissue_type="Whole seedling, 4 day old"
/lab_host="DH10B"
/clone_lib="Gm-cl056"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from 4 day old seedling of P1468916. The seedlings were germinated in a growth chamber using germination paper. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

ORIGIN

Query Match 80.0%; Score 16.8; DB 14; Length 325;
Best Local Similarity 90.0%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCT 20
||||| |||||||
Db 181 AGGAGCGCGCAAAACCTTTT 200

RESULT 7

CG905629/c 379 bp DNA linear GSS 09-DEC-2003
LOCUS ZMMBB0515F10r ZMMBBB (HindIII) Zea mays subsp. mays genomic clone
DEFINITION ZMMBB0515F10 3', genomic survey sequence.
ACCESSION CG905629
VERSION CG905629.1 GI:39604896
KEYWORDS GSS.
SOURCE Zea mays subsp. mays (maize)

ORGANISM

Zea mays subsp. mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 379)

Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,

Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.

Sequencing of the maize genome at PGIR (2003c)

Unpublished (2003)

Contact: Bharti,A.K.

Dr.Joachim Messing's lab

The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers University

190 Frelinghuysen Road, Piscataway, NJ 08854, USA

Tel: 732 445 3801

Fax: 732 445 5735

Email: bharti@waksman.rutgers.edu

Seq primer: SP6

Class: BAC ends
High quality sequence start: 113.
Location/Qualifiers
1. .379

/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBB0515F10"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBB (HindIII)"
/note="Vector: pCUGI; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 80.0%; Score 16.8; DB 29; Length 379;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCT 20
||||| |||||||
Db 185 AGGAGCGCGCAAAAGCTTCT 166

RESULT 8

CG701116 409 bp DNA linear GSS 15-OCT-2003
LOCUS ZMMBB0117D10r ZMMBBB (EcoRI) Zea mays subsp. mays genomic clone
DEFINITION ZMMBB0117D10 3', genomic survey sequence.
ACCESSION CG701116
VERSION CG701116.1 GI:37688917
KEYWORDS GSS.

SOURCE Zea mays subsp. mays (maize)

ORGANISM

Zea mays subsp. mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 409)

Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,

Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.

Sequencing of the maize genome at PGIR (2003c)

Unpublished (2003)

Contact: Bharti,A.K.

Dr.Joachim Messing's lab

The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers University

190 Frelinghuysen Road, Piscataway, NJ 08854, USA

Tel: 732 445 3801

Fax: 732 445 5735

Email: bharti@waksman.rutgers.edu

Seq primer: SP6

Class: BAC ends

High quality sequence start: 137.

Location/Qualifiers

FEATURES

source

1. .409
Location/Qualifiers
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBB0117D10"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBB (EcoRI)"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN

Query Match 80.0%; Score 16.8; DB 29; Length 409;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCT 20
||||| |||||||
Db 166 AGGAGCGCGCAAAACCTTCT 185

```

RESULT 9
LOCUS CG878712/c
DEFINITION ZMMBBb0499A04r ZMMBBB (HindIII) Zea mays subsp. mays genomic clone
ACCESSION CG878712
VERSION CG878712.1 GI:38609315
KEYWORDS GSS.
SOURCE Zea mays subsp. mays (maize)
ORGANISM Zea mays subsp. mays
REFERENCE 1 (bases 1 to 427)
AUTHORS Bhatti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Zohovetz,V., Fuks,G., Yu,Y., Wang,R. and Messing,J.
TITLE Sequencing of the maize genome at PGIR (2003c)
JOURNAL Unpublished (2003)
COMMENT Contact: Bhatti,A.K.
Dr Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bhatti@waksman.rutgers.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 111.
Location/Qualifiers
1..427
/mol_type="genomic DNA"
/organism="Zea mays subsp. mays"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBBb0499A04"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBB (HindIII)"
/notes="Vector: pCUGI; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match 80.0%; Score 16.8; DB 29; Length 427;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTCTT 20
DB 197 AGGAGCGTGCAGAAACCTCTT 178

RESULT 10
LOCUS AA817025
DEFINITION LD22010.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
melanogaster cDNA clone LD22010 5 similar to CG11583: Fban0011583
GO: (l) located on: 3L 64B6-64B6;; 08/13/2002, mRNA sequence.
ACCESSION AA817025
VERSION AA817025.1 GI:2886634
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 443)
AUTHORS Harvey,D., Brockstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: LD22010.3prime

us-10-071-411a-4.rst
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003480; arm:3L [3969219,4279711]
estimated-cyto:64A2-64B6; 04/16/2001
Plate: LD.220 row: A column: 10
High quality sequence stop: 336
POLYA=No.

FEATURES
source
1..443
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LD22010"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/notes="Organ: embryo; Vector: pOT2; Site 1: EcoRI; Site 2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. "

ORIGIN
Query Match 80.0%; Score 16.8; DB 9; Length 443;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGAGCGCGCAAAACCTCTC 21
DB 312 GGAGCGCTCAAAACCTTGT 331

RESULT 11
LOCUS AA739254
DEFINITION V51e01.r1 Soares thymus_2NBMT Mus musculus cDNA clone
IMAGE:1225944 5, similar to gb:Y00769 Murine mRNA for integrin beta
subunit (MOUSE);, mRNA sequence.
ACCESSION AA739254
VERSION AA739254.1 GI:2775440
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 446)
REFERENCE 1
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
JOURNAL JOURNAL
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:651536
Putative full length read
vector to vector length is
Seq primer: -28mi3 rev2 Et from Amersham
High quality sequence stop: 322.
Location/Qualifiers
1..446
/organism="Mus musculus"

FEATURES
source
1..446
/organism="Mus musculus"

```



```

DEFINITION BOGQP74TR BOGQ Brassica oleracea genomic clone BOGQP74, genomic
survey sequence.
ACCESSION BH486134
VERSION BH486134.1 GI:117694238
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
JOURNAL 1 (bases 1 to 584)
COMMENT Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOGQP74TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES             Location/Qualifiers
     source            1..584
                        /organism="Brassica oleracea"
                        /mol_type="genomic DNA"
                        /strain="TO1000DH3"
                        /db_xref="taxon:3712"
                        /clone_lib="BOGQP74"
                        /notes="BOGQP74"
ORIGIN
Query Match          80.0%; Score 16.8; DB 28; Length 584;
Best Local Similarity 90.0%; Pred. No. 1.7e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  GGAGCGCGCAAAACCTTCTC 21
      ||||| ||||| ||||| |||||
DB      557  GCAGCGCGCAAAACCTTCTC 538

Query Match          80.0%; Score 16.8; DB 28; Length 584;
Best Local Similarity 90.0%; Pred. No. 1.7e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  GGAGCGCGCAAAACCTTCTC 21
      ||||| ||||| ||||| |||||
DB      557  GCAGCGCGCAAAACCTTCTC 538

Search completed: May 7, 2004, 11:54:21
Job time : 1443.5 secs

FEATURES             Location/Qualifiers
     source            1..608
                        /organism="Aedes aegypti"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7159"
                        /clone_lib="CSU-K34.120B1"
                        /clone_lib="CSU-K34"
                        /note="Vector: pBACe3.6; Site 1: EcoRI; Source DNA: Aedes
aegypti; strain unknown [derived from freshly hatched
larvae at the Virus Research Centre, Poona, India.
Reference: SINGH, K. R. P., 1967 Cell cultures derived
from larvae of Aedes albopictus (Skuse) and Aedes aegypti
(L.). Current Science 36: 506-508]; ATC-10 cell line ATCC
CCL-125"
ORIGIN
Query Match          80.0%; Score 16.8; DB 28; Length 608;
Best Local Similarity 90.0%; Pred. No. 1.7e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  GGAGCGCGCAAAACCTTCTC 21
      ||||| ||||| ||||| |||||
DB      483  GGAGCGCTCACACCTTCTC 464

Search completed: May 7, 2004, 11:54:21
Job time : 1443.5 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 09:53:47 ; Search time 667.597 Seconds
(without alignments)
1363.403 Million cell updates/sec

Title: US-10-071-411a-4

Perfect score: 21

Sequence: 1 aggagcgcgaacattctc 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_ini:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pi:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19.4	92.4	2189	6	AR199380	AR199380 Sequence
2	19.4	92.4	2189	6	AX285281	AX285281 Sequence
3	19.4	92.4	2189	9	HUMLIF05	M38191 Human 5-lip
4	19.4	92.4	129266	9	AL731567	AL731567 Human DNA
5	19.4	92.4	160654	2	AC011879	AC011879 Homo sapi
6	17.8	84.8	2361	1	TSDNALIG	Z29528 T.scotoduct
7	17.8	84.8	261936	2	AC103180	AC103180 Rattus no
8	17.4	82.9	331801	1	NMA422491	AL162755 Neisseria
9	16.8	80.0	8472	1	AE008076	AE008076 Agrobacte
10	16.8	80.0	10579	1	AE002431	AE002431 Neisseria
11	16.8	80.0	13051	1	AE009111	AE009111 Agrobacte
12	16.8	80.0	173503	2	BX890641	BX890641 Danio rer
13	16.8	80.0	228821	2	AC098626	AC098626 Rattus no
14	16.8	80.0	231028	2	AC105707	AC105707 Rattus no
15	16.8	80.0	240116	2	BX248332	BX248332 Danio rer
16	16.8	80.0	242445	2	AC105670	AC105670 Rattus no
17	16.8	80.0	244180	2	AC147558	AC147558 Mus muscu
18	16.8	80.0	245178	2	AC125747	AC125747 Rattus no
19	16.8	80.0	245257	2	AC135577	AC135577 Rattus no
20	16.8	80.0	253261	2	AC127737	AC127737 Rattus no
21	16.8	80.0	261493	2	AC106564	AC106564 Rattus no
22	16.8	80.0	267759	2	AC079418	AC079418 Mus muscu
23	16.8	80.0	287626	2	AC111963	AC111963 Rattus no
24	16.8	80.0	349980	6	AX044030	AX044030 Sequence
25	16.4	78.1	890	8	AMRBCSM3	X51813 Acetabulari
26	16.4	78.1	1463	1	AY345543	AY345543 Unidentif
27	16.4	78.1	145295	5	BX005348	BX005348 Zebrafish
28	16.4	78.1	155323	2	BX511299	BX511299 Danio rer
29	16.4	78.1	205411	2	BX119920	BX119920 Danio rer
30	16.4	78.1	229278	2	RX3233548	RX3233548 Danio rer
31	16.2	77.1	627	4	AY356121	AY356121 Canis fam
32	16.2	77.1	1350	1	AY227046	AY227046 Rhizobium
33	16.2	77.1	1441	8	ZMU17979	U17979 Zea mays tr
34	16.2	77.1	1515	1	AF092867	AF092867 Thermus s
35	16.2	77.1	1638	1	AF092865	AF092865 Thermus f
36	16.2	77.1	1671	1	AF092864	AF092864 Thermus a
37	16.2	77.1	2051	6	AR051693	AR051693 Sequence
38	16.2	77.1	2051	6	I18226	I18226 Sequence 7
39	16.2	77.1	2051	6	BD079297	BD079297 DNA ampli
40	16.2	77.1	2100	1	TTHDNALGS	M74792 Thermus the
41	16.2	77.1	2111	6	AR051690	AR051690 Sequence
42	16.2	77.1	2111	6	I18222	I18222 Sequence 1
43	16.2	77.1	2111	6	BD079293	BD079293 DNA ampli
44	16.2	77.1	2436	1	TTHDNALIG	M36417 Thermus the
45	16.2	77.1	15705	3	DMEN32G11	AL035632 Drosophil

ALIGNMENTS

RESULT 1
AR199380
LOCUS AR199380 2189 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355434.
ACCESSION AR199380
VERSION AR199380.1 GI:20249454
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2189)
AUTHORS Drazen,J.M., In,K.-H., Asano,K., Beier,D. and Grobholz,J.
TITLE 5-Lipoxygenase Gene polymorphisms and their use in classifying patients
JOURNAL Patent: US 6355434-A 1 12-MAR-2002;

FEATURES
source
Location/Qualifiers
1..2189
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 92.4%; Score 19.4; DB 6; Length 2189;
Best Local Similarity 95.2%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGAGCGCGCAAAACCTTCTC 21
|||||
990 AGGAGCGCGCAAAACCTTCTC 1010

Db

RESULT 2
AX285281 2189 bp DNA linear PAT 20-NOV-2001
LOCUS
DEFINITION Sequence 5 from Patent WO0179560.
ACCESSION AX285281
VERSION AX285281.1 GI:17045963
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Anderson, W.H., Edwards, L.D., Emmett, A.H., Pillai, S. and
Sprankel, C.S.
TITLE Medicine response assay in respiratory disease
JOURNAL Patent: WO 0179560-A 5 25-OCT-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
source
Location/Qualifiers
1..2189
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 92.4%; Score 19.4; DB 6; Length 2189;
Best Local Similarity 95.2%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGAGCGCGCAAAACCTTCTC 21
|||||
990 AGGAGCGCGCAAAACCTTCTC 1010

Db

RESULT 3
HUMLIPO5 2189 bp DNA linear PRI 07-MAR-1995
LOCUS
DEFINITION Human 5-lipoxygenase gene, exon 1.
ACCESSION M38191
VERSION M38191.1 GI:187166
KEYWORDS 5-lipoxygenase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2189)
Hoshiko, S., Radmark, O. and Samuelsson, B.
TITLE Characterization of the human 5-lipoxygenase gene promoter
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (23), 9073-9077 (1990)
MEDLINE 91067649
FURNISHED 2251250
COMMENT Original source text: Human DNA, clone lx12A.
Draft entry and computer-readable sequence for [Proc. Natl. Acad. Sci. U.S.A. (1990) In press] kindly submitted by O. Radmark, 28-AUG-1990.
FEATURES
source
Location/Qualifiers
1..2189
/organism="Homo sapiens"
/mol_type="genomic DNA"

FEATURES
source
Location/Qualifiers
1..2189
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 92.4%; Score 19.4; DB 9; Length 2189;
Best Local Similarity 95.2%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGAGCGCGCAAAACCTTCTC 21
|||||
990 AGGAGCGCGCAAAACCTTCTC 1010

Db

RESULT 4
AL731567 129266 bp DNA linear PRI 20-JUN-2002
LOCUS
DEFINITION Human DNA sequence from clone Rp11-67C2 on chromosome 10, complete sequence.
ACCESSION AL731567 AC010865
VERSION AL731567.6 GI:21537524
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 129266)
Whitehead, S.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jun 21, 2002 this sequence version replaced gi:212133582.
Draft Sequence Produced by Genome Therapeutics Corp, 100 Beaver Street, Waltham, MA 02453, USA
http://www.genomecorp.com
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr10>
 RP11-67C2 is from the library RPCI-11.1 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6.

FEATURES

Location/Qualifiers
 1. 129266
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-67C2"
 /clone_lib="RPCI-11.1"

ORIGIN

Query Match 92.4%; Score 19.4; DB 9; Length 129266;
 Best Local Similarity 95.2%; Pred No. 65;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 AGGAGCGCGCAAAACCTTCTC 21
 |||||
 Db 33726 AGGAGCGCGCAAAACCTTCTC 33746

RESULT 5

AC011879 160654 bp DNA linear HTG 16-MAR-2000
 LOCUS Homo sapiens clone RP11-16P14, WORKING DRAFT SEQUENCE, 30 unordered
 DEFINITION pieces.

ACCESSION

AC011879
 VERSION AC011879.3 GI:7239554
 KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.

SOURCE

SOURCE Homo sapiens (human)

ORGANISM

Organism: Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 160654)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens, clone RP11-16P14

Unpublished

2 (bases 1 to 160654)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckhalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,

Cooke, P., DeRellano, K., Dewar, K., Domino, M., Donnelly, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J., Johnson, R., Jones, C., Kamm, L., Karatas, A., Klein, J.,

Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 14, 2000 this sequence version replaced gi:6524208.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIGR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submission@genome.wi.mit.edu

----- Project Information

Center project name: L3606

Center clone name: 16.P.14

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
 Consensus quality: 111055 bases at least Q40
 Consensus quality: 135086 bases at least Q30
 Consensus quality: 147921 bases at least Q20
 Insert size: 163000; agarose-fp
 Quality coverage: 2.9 in Q20 bases; agarose-fp
 Quality coverage: 3.0 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 consists of 30 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 151: contig of 151 bp in length
 152: gap of 100 bp
 252 1760: contig of 1509 bp in length
 1761 1860: gap of 100 bp
 1861 3069: contig of 1209 bp in length
 3070 3169: gap of 100 bp
 3170 4720: contig of 1551 bp in length
 4721 4820: gap of 100 bp
 4821 6174: contig of 1354 bp in length
 6175 6274: gap of 100 bp
 6275 7417: contig of 1143 bp in length
 7418 7517: gap of 100 bp
 7518 9158: contig of 1641 bp in length
 9159 9258: gap of 100 bp
 9259 10865: contig of 1607 bp in length
 10866 10965: gap of 100 bp
 10966 12859: contig of 1894 bp in length
 12860 12959: gap of 100 bp
 12960 15671: contig of 2712 bp in length
 15672 15771: gap of 100 bp
 15772 18082: contig of 2311 bp in length
 18083 18182: gap of 100 bp
 18183 20523: contig of 2341 bp in length
 20524 20623: gap of 100 bp
 20624 22903: contig of 2280 bp in length
 22904 23003: gap of 100 bp
 23004 23671: contig of 668 bp in length
 23672 23771: gap of 100 bp
 23772 25541: contig of 1770 bp in length
 25542 25641: gap of 100 bp
 25642 28323: contig of 2682 bp in length
 28324 28423: gap of 100 bp
 28424 31498: contig of 3075 bp in length
 31499 31598: gap of 100 bp
 31599 36626: contig of 5028 bp in length
 36627 36726: gap of 100 bp
 36727 42109: contig of 5383 bp in length
 42110 42209: gap of 100 bp
 42210 48339: contig of 6130 bp in length
 48340 48439: gap of 100 bp
 48440 55333: contig of 6894 bp in length
 55334 55433: gap of 100 bp
 55434 63595: contig of 8161 bp in length
 63596 63694: gap of 100 bp
 63695 73943: contig of 10249 bp in length
 73944 74043: gap of 100 bp
 74044 83665: contig of 9622 bp in length
 83666 83765: gap of 100 bp
 83766 95324: contig of 11557 bp in length
 95325 95423: gap of 100 bp
 95424 108403: contig of 12981 bp in length
 108404 108503: gap of 100 bp
 108504 120220: contig of 11717 bp in length
 120221 120320: gap of 100 bp
 120321 132958: contig of 12638 bp in length
 132959 133058: gap of 100 bp

```

* 133059 145697: contig of 12639 bp in length
* 145698 145797: gap of 100 bp
* 145798 160654: contig of 14857 bp in length.
FEATURES
  source
    1..160654
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone="RP11-16P14"
      /clone_lib="RPC1-11 Human Male BAC"
    1..151
      /note="assembly_fragment"
      clone_end:17
      vector_side:right
    252..1760
      /note="assembly_fragment"
    1861..3069
      /note="assembly_fragment"
    3170..4720
      /note="assembly_fragment"
    4821..6174
      /note="assembly_fragment"
    6275..7417
      /note="assembly_fragment"
    7518..9158
      /note="assembly_fragment"
    9259..10865
      /note="assembly_fragment"
    10966..12859
      /note="assembly_fragment"
    12960..15671
      /note="assembly_fragment"
    15772..18082
      /note="assembly_fragment"
    18183..20523
      /note="assembly_fragment"
    20624..22903
      /note="assembly_fragment"
    23004..23671
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:right
    23772..25541
      /note="assembly_fragment"
    25642..28323
      /note="assembly_fragment"
    28424..31498
      /note="assembly_fragment"
    31599..36626
      /note="assembly_fragment"
    36727..42109
      /note="assembly_fragment"
    42210..48339
      /note="assembly_fragment"
    48440..55333
      /note="assembly_fragment"
    55434..63594
      /note="assembly_fragment"
    63695..73943
      /note="assembly_fragment"
    74044..83665
      /note="assembly_fragment"
    83766..95322
      /note="assembly_fragment"
    95423..108403
      /note="assembly_fragment"
    108504..120220
      /note="assembly_fragment"
    120321..132958
      /note="assembly_fragment"
    133059..145697
      /note="assembly_fragment"
    145798..160654

```

```

ORIGIN
Query Match          92.4%; Score 19.4; DB 2; Length 160654;
Best Local Similarity 95.2%; Pred. No. 64;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCTC 21
    |||||
Db 40803 AGGAGCGCGCAAAACCTTCTC 40823

RESULT 6
TSDNALIG          2361 bp DNA linear BCT 20-APR-1995
T.scotoductus (DSM 8553) gene for DNA ligase.
ACCESSION Z29528
VERSION Z29528.1 GI:609275
KEYWORDS DNA ligase.
SOURCE Thermus scotoductus
ORGANISM Thermus scotoductus
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE 1 (bases 1 to 2361)
AUTHORS Jonsson,Z.O., Thorbjarnardottir,S.H., Eggertsson,G. and
Palsdottir,A.
TITLE Sequence of the DNA ligase-encoding gene from Thermus scotoductus
and conserved motifs in DNA ligases
JOURNAL Gene 151 (1-2), 177-180 (1994)
MEDLINE 95129855
PUBMED 7828870
REFERENCE 2 (bases 1 to 2361)
AUTHORS Jonsson,Z.O.
TITLE Direct Submision
JOURNAL Submitted (18-JAN-1994) Zophonias O. Jonsson, Department of
Biology, University of, Iceland, Grensasvegur 12, Reykjavik,
IS-108, Iceland
FEATURES
  source
    1..2361
      /organism="Thermus scotoductus"
      /mol_type="genomic DNA"
      /strain="DSM 8553"
      /db_xref="taxon:37636"
      /clone="pHIS5Ra"
      /EC_number="6.5.1.2"
      /codon_start=1
      /transl_table=11
      /product="DNA ligase"
      /protein_id="CAA82645.1"
      /db_xref="GI:609276"
      /db_xref="GOA:P49422"
      /translation="MTLEARKVNLRLIRYHYVVLADPEISDAEYDRLREL
      KELEERFPELKSPTSPTQVGAKPLEATFRPIRHPTRMYSLDNAFNPDELKAFERIG
      RALRGEPFAYTVEHKVGLSVNLVYEDGVLVWGTRGDGEGVETONLITPIPR
      RVKGPVERLEVRGEVYMPTEAPFLNLEELKEGKIFKNPRNAAAGSRQKDPRIAR
      RGLRATFYALGLGESEGLTQLDLHLWLRKGFVFERHGFARAEAGVRYIQQWLK
      TGRVTPVGLIEPVTEGVSVRVTHNESYIEELDVRIGDWLVHKGAVIPELVRLV
      KEKRTGERPPIRWPETCPCCGRLVKEGVHRCNPCLCAKFEAIRHVASRKAMDIG
      GLGEKLIKLEGLVADLYRLKKDDLLGLRMGEKSAQNLRLQIEESKRGCLER
      LLYALGLFVGEVLAARNAHFMTDRLLLEASLELLQVEVEGELTARGIYETLQDPA
      FRDLVRLKEAGVVMKEARGEALKGLTFVITGELSPREVEKALLRRLKAKVTDVS
      SRKTSYLVVGENPGSKLEKARALGVPTLTTEBELYRLIEERTCKPVETLAS"

```

ORIGIN

```

Query Match          84.8%; Score 17.8; DB 1; Length 2361;
Best Local Similarity 90.5%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 AGGAGCGCGCAAAACCTTCTC 21
    |||||

```

Db 1777 AAGAGCGCGCAAAACCTCTC 1797

RESULT 7
AC103180/c
LOCUS
DEFINITION AC103180 261936 bp DNA linear HTG 13-MAY-2003
*** 3 unordered pieces.
AC103180
VERSION AC103180.6 GI:30578543
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 261936)
Muzny,D,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Kocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackschall,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowals,C., Kraft,C.I., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenstehew,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naik,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reich,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villabana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Woodson,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 261936)
Worley,K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE JOURNAL
AUTHORS
TITLE
JOURNAL

Query Match 84.8%; Score 17.8; DB 2; Length 261936;
Best Local Similarity 90.5%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

REFERENCE JOURNAL
AUTHORS
TITLE
JOURNAL

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 261936)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24819226.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJJK
Center clone name: CH230-10113
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 232513 bases at least Q40
Consensus quality: 236922 bases at least Q30
Consensus quality: 240613 bases at least Q20
Estimated insert size: 245719; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 257409: contig of 257409 bp in length
* 257410 257509: gap of unknown length
* 257510 258900: contig of 1391 bp in length
* 258901 259000: gap of unknown length
* 259001 261936: contig of 2936 bp in length.

FEATURES
source
1..261936
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-10113"
28..495
/note="clone boundary
clone end:T7
site:EcoRI
end_sequence:BH312072"
complement(255406..255853)
/note="clone boundary
clone end:Sp6
site:EcoRI
end_sequence:BH312074"

misc_feature
28..495
/note="clone boundary
clone end:T7
site:EcoRI
end_sequence:BH312072"
complement(255406..255853)
/note="clone boundary
clone end:Sp6
site:EcoRI
end_sequence:BH312074"

misc_feature
28..495
/note="clone boundary
clone end:T7
site:EcoRI
end_sequence:BH312072"
complement(255406..255853)
/note="clone boundary
clone end:Sp6
site:EcoRI
end_sequence:BH312074"

ORIGIN
Query Match 84.8%; Score 17.8; DB 2; Length 261936;
Best Local Similarity 90.5%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy      1 AGGAGCGCGCAAAACCTTCTC 21
      |||||
Db      18692 AGGAGCGCGCAAAACCTTCCC 18672

RESULT 8
LOCUS   NMA422491          331801 bp      DNA      linear      BCT 02-SEP-2002
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 4/7.
ACCESSION   AL162755
VERSION     AL162755.2  GI:7379742
KEYWORDS
SOURCE
ORGANISM   Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1 (bases 1 to 331801)
AUTHORS    Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
            Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
            Nature 404 (6777), 502-506 (2000)
JOURNAL    Nature
MEDLINE    20222556
PubMed     10761919
REFERENCE   2 (bases 1 to 331801)
AUTHORS    Parkhill,J.
TITLE      Direct Submission
JOURNAL    Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N\_meningitidis/).
FEATURES             Location/Qualifiers
     source            1..331801
                        /organism="Neisseria meningitidis Z2491"
                        /mol_type="genomic DNA"
                        /strain="Z2491"
                        /db_xref="taxon:122587"
     misc_feature      117..1333
                        /note="serogroup: A"
     gene              117..1333
                        /gene="NMA1040"
                        /pseudo
     CDS                117..1333
                        /gene="NMA1040"
                        /notes="NMA1040, possible type I restriction-modification
                        system specificity protein, pseudogene, len: 1217 bp;
                        N-terminus shows weak similarity to the C-terminal half of
                        TR:Q50359 (EMBL:U25415), hsdS1B, Mycoplasma pulmonis
                        restriction-modification enzyme subunit S1B (336 aa),
                        fasta scores; E(): 2.3e-06, 26.2% identity in 141 aa
                        overlap and to many hypothetical restriction-modification
                        subunits. Also similar to NMA1041, fasta scores; E():
                        4.8e-10, 37.0% identity in 100 aa overlap. C-terminus
                        similar to part of SW:TL1_ECOLI (EMBL:X13145), hsdS
                        Escherichia coli type I restriction enzyme EcoR124II
                        specificity protein (410 aa), fasta scores; E(): 1.9e-14,
                        39.6% identity in 182 aa overlap. Similar to NMA1040, E():
                        2.9e-09, 38.0% identity in 100 aa overlap. Contains Pfam
                        match to entry PF01420 Methylase S, Type I restriction
                        modification DNA specificity domain. Contains a G(8) tract
                        which would allow translation as an intact CDS, if
                        variable. Lies within a region of unusually low GC
                        content"
                        /pseudo
                        /codon_start=1
                        /transl_table=11

```

```

/misc_feature
/misc_feature
repeat_region
/misc_feature
/misc_feature
misc_feature
RBS
gene
CDS
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
stem_loop
misc_feature
misc_feature
gene

```

```

/product="pseudogene (putative type I
restriction-modification system specificity protein)"
complement(157..166)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
171..1620
/gene="NMA1040"
/note="Pfam match to entry PF01420 Methylase S, Type I
restriction modification DNA specificity domain, score
105.50, E-value 1.1e-27"
/pseudo
699..706
/note="(g)8"
726..735
/gene="NMA1040"
/note="Core DNA uptake sequence: gccgtctgaa"
/pseudo
/label=DUS
740..1231
/gene="NMA1040"
/note="Pfam match to entry PF01420 Methylase S, Type I
restriction modification DNA specificity domain, score
158.70, E-value 1e-43"
/pseudo
1405..1408
1415..4509
/gene="NMA1042"
/pseudo
1415..4509
/gene="NMA1042"
/note="NMA1042, pseudogene, probable type I
restriction-modification system restriction protein, len:
291 aa; similar to parts of many e.g. SW:TL1_ECOLI
(EMBL:X13145), hsdR, Escherichia coli type I restriction
enzyme EcoR124II R protein (EC 3.1.21.3) (1033 aa), fasta
scores; E(): 0.73, 9% identity in 291 aa overlap (frame
1), followed by a stop codon, E(): 0.92, 9% identity in
127 aa overlap (frame 1), followed by a frameshift near a
poly-A tract, E(): 0.72, 5% identity in 604 aa overlap
(frame 3)"
/pseudo
/codon_start=1
/transl_table=11
/product="type I restriction-modification system
restriction protein (pseudogene)"
complement(1642..1651)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1687..1696
/gene="NMA1042"
/note="Core DNA uptake sequence: gccgtctgaa"
/pseudo
/label=DUS
1761..1770
/gene="NMA1042"
/note="Core DNA uptake sequence: gccgtctgaa"
/pseudo
3906..3915
/gene="NMA1042"
/note="Core DNA uptake sequence: gccgtctgaa"
/pseudo
/label=DUS
4514..4545
/note="stem loop containing DNA uptake sequences: aaaat
gccgtctgaa at ttgagagcc atttt"
4519..4528
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(4531..4540)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(4559..4638)

```



```
/gene="clpA"
complement(4559..6838)
/gene="clpA"
note="NMA1045, clpA, probable ATP-dependent protease
ATP-binding protein, len: 759 aa; similar to many e.g.
SW:CLPA_ECOLI (EMBL:M31045), clpA, Escherichia coli
ATP-dependent Clp protease ATP-binding subunit (758 aa),
fasta scores: E(): 0, 56.2% identity in 762 aa overlap.
Similar to NMA1683, fasta scores: E(): 0, 37.2% identity
in 844 aa overlap. Contains Pfam match to entry PF00495
clpA_B, Chaperonin clpA/B, PS00870 Chaperonins clpA/B
signature 1, PS00871 Chaperonins clpA/B signature 2 and
two PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent protease ATP-binding
protein"
/protein_id="CAB84311.1"
/db_xref="GI:7379743"
/db_xref="GOA:O9JVL5"
/db_xref="SPTREMBL:O9JVL5"
/translation="MLSPELEHLLQLLYREARNARYEPTISLEHLLVLEEDAAVNV
LKLGDADLVSEQLAASVAENTPOLPDHLLDTVETQPLGFRQVIRAWHTQAGK
AAVEPLDILVAMWSESHAVYFLKLSQSTRFEVLCRIAHGSPDDGNDGLGREG
EAEQKTSLSYTNLNAEVKAGRIDPLIGRKHEMERLIVQILCRKNNPLLVGEAG
VCKTALAELGHOIVKGDIPDPAKDAEVVALDMSGLLACTKVRGFEARVKSVLQLE
KIPHALFTDEHTIIGAGTSGTMDASNLKPKALGKRCIGATTVEYRTYFDK
DHLSRRFKIDVVEFTVSETVQILRGLKPMEGFHQVRYTQGALEAAELSAKYNE
REFLPKADVDEAGAKQILPKSKOKKVIKGAQETVIKAVRIPEKTVSHDDKQVL
QPLGRDLKNMVGQENAIYALVAAMKSRSGLGPKDPTIGSFISFGPTGVGTQVAKQ
LAYSMGVPLQFDMGSEMERHAYVRLIGAPGVGVFEQGLLTERAVNKQPHCVLLDE
IKAHPIENVLVMDAGKLTDDNGKSADFRNVLIMTNAGASLSRPSLGFTAKR
FEGDEMOANKLFTPEFRNLDAIIPFAPLSEPIIVKVVDFKLLQLEHLLDKKVEAE
FPALKYLAEGFDPOMGARNPLRQEKIRKPLADELLFGKLVDDGGFVDDWDAK
BEAVLKFKSKVKKPTETV"
complement(4637..5578)
/gene="clpA"
/notes="Pfam match to entry PF00495 clpA_B, Chaperonin
clpA/B, score 583.70, E-value 1.1e-171"
complement(5228..5284)
/gene="clpA"
/notes="PS00871 Chaperonins clpA/B signature 2"
complement(5330..5353)
/gene="clpA"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(5894..5932)
/gene="clpA"
/notes="PS00870 Chaperonins clpA/B signature 1"
complement(6173..6196)
/gene="clpA"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
6374..6383
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(6842..7153)
/gene="NMA1046"
complement(6842..7153)
/gene="NMA1046"
/notes="NMA1046, len: 103 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YLJA_ECOLI (EMBL:AE000190),
yljA, Escherichia coli hypothetical protein (103 aa),
Query Match 82.9%; Score 17.4; DB 1; Length 331801;
Best Local Similarity 94.7%; Pred. No. 6.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 2 GGAGCGCGCAAAACCTTCT 20
Db 148760 GAAGCGCGCAAAACCTTCT 148778
RESULT 9
AE008076
```

AE008076 8472 bp DNA linear BCT 18-DEC-2001
Agrobacterium tumefaciens str. C58 circular chromosome, section 134
of 254 of the complete sequence.
AE008076 AE007869
AE008076.1 GI:15156608
Agrobacterium tumefaciens str. C58 (Cereon)
Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
1 (bases 1 to 8472)
Hinkle, G., Slater, S.C. and Goodner, B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
2 (bases 1 to 8472)
Hinkle, G., Slater, S.C. and Goodner, B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
Location/Qualifiers
1. 8472
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/mol_type="genomic DNA"
/strain="C58"
/db_xref="taxon:181661"
79..1077
/gene="AGR_C_2805"
79..1077
/gene="AGR_C_2805"
/note="(AE005408) putative adhesin"
/codon_start=1
/transl_table=11
/product="AGR_C_2805p"
/protein_id="AAK87312.1"
/db_xref="GI:15156609"
/translation="MHFMKSLIPLMASAALAAVAGATAADVDVVVSIKPVHSLVAAL
MKGVEPOLIVDGAASPHYTLNRPNSARKLEKADVFWVGPGLAEFLQKPLALASKA
TVLELDAKLEKLPFRKGGPEAHDDGEGEHAHAGHTEDEGHDHNDHAGSEHE
HGAYDTHLWLDPANAKAAQAETALIAADAGNAATYQANTKLLDLDLDAEVVET
VVPKDKPFLVEHDAYQVFEHYGVKTAGSITVSPETLPADRVKQMKVQLGAPC
VFAPQEPFKLVSVITGTAASKATLDPAAATLTPGPDLYFLKMRGIAGSLKNCLS"
complement(1099..2166)
/gene="AGR_C_2807"
complement(1099..2166)
/gene="AGR_C_2807"
/note="TRANSCRIPTION REGULATOR AGLR"
/codon_start=1
/transl_table=11
/product="AGR_C_2807p"
/protein_id="AAK87313.1"
/db_xref="GI:15156610"
/translation="MWRNDMDTKSGGDMPSAMGCRPLTKTIAYMTGLGITVSRM
LKDPDGAETKERVLLIARQIQYQNRAGVRLRGKTNVIALVLSDELMEGFTSQM
VFGITEVLTSTQTHLVVTHPIKAKOSNVPRIYLETGSDAGVLIISKLENDPRVEMT
ERNMPVTHGRSDMGIEHAFHDFDNEAYAEVAERLAQCGRKRIAVIPPSRFSFDH
ARKGNRGRDIFGLTEFFDAVTIETPLEKIRDFGQRLMQSDRPDGIIVISGSSTIA
LVAGPEAAQGVKIGEDVDIVSKQSAEFLNMIKPIQHTVNEDIKLAGRELAKALLARNG
LVAETIQLSGSPWSSMAKP"
2640..5138
/gene="AGR_C_2809"
2640..5138
/gene="AGR_C_2809"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2809p"
/protein_id="AAK87314.1"
/db_xref="GI:15156611"
/translation="MQGPIWISSSTPETVIDLAGLWHLASVGEHGHAETISIPGDIHSA
LKNAILPDPYHGANEKAVQWVAQDWILERTILDDAEASWYLDLDYLDVIAIVFN
DVPVLADNCFRRYRDPDISRAVRGENTIRHFHSNTAGAEARQAPFYIPYHSGNS


```

/db_xref="GI:7225999"
/translation="MTAKEELFAWLRHMQNKGSLFVTHPPPAKMDGKILRITDE
PLTAKCMEIAFSINSAQOAEFFSTNECFALSDPDSFRFVNAMIQGKALVFRIT
ITSIKPFESLNPVLKDVALKKGLVIFVGTGSGKSTSLASLDIYRNSFGHII
TIEDIEFVHKNKCIITQREVGVDTENMAALKNLRQAPDVLIGIRDETMDYA
LAFATGHLCAWTLHANGTNOALDRIINFFPEEREOQLTOLSNLOAFISORLVPRD
GSGKVAEVLNLSLISELHNGNIHEIKVEMKSKTTLGWTQTPDQHLQLYEKGDI
SLOEALKNADSHDLRLAVQLRSRAQSSSPDELL"
2136..3113
/gene="NMB0769"
CDS
2136..3113
/gene="NMB0769"
/notes="similar to GP:145799 percent identity: 50.39;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, delta prime subunit,
putative"
/protein_id="AAF41182.1"
/db_xref="GI:7226000"
/translation="MIYPWHNQWQIAEHWRERRNMLFAGKDKTGKTFARPAKA
LTCETPACPKCGSCMSCHLFGQSHDPFYETITLSDPEPNRKLQIKIDAVREII
DNVYITSVRGGLRVLIHPSAEMVQAANSLKLVLEPPQVQVFLVSHADKVLPTI
KSRCKWVLPAPSHAEALAYLRERGVAEPERLAFHSGAPLDEADGVRLRIKLIDI
LAEPRLKTLIDVAALFDKELPLAVFGVMQKLVLDGLCLQHMKPVVYPAYEDRLIQ
TVSGRPNRVFAEDMLKQLAPYGFHTLVNKVQIEHLLINYLELKKENG"
3117..3467
/gene="NMB0770"
CDS
3117..3467
/gene="NMB0770"
/notes="similar to PID:972779 percent identity: 62.62;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="type IV pilus assembly protein PilZ, putative"
/protein_id="AAF41183.1"
/db_xref="GI:7226001"
/translation="WSDGONIPAKWMSLQKDMNLLYSSYMPFLEHGGFLVOTNDVFS
IGDILLAVEILNFKPLFTTKVWINPARTSKPKGVGLATKHCNCLKVKDQLEVE
LGTIGGSRPTFTM"
3472..4251
/gene="NMB0771"
CDS
3472..4251
/gene="NMB0771"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF41184.1"
/db_xref="GI:7226002"
/translation="MHIIIDSHCHLFEGLKERLPEVLSNNEANGVQALAIYSVRESF
SEVFAIAEAHEHYCTIGVHPDSKEAEBSIAEMVEAHKPKVGVGEGGLDYCKG
DLSQHKRFADHAEANQGLPVIVHTRDAADTLSEKCRVNSGVTHCFSEDIGFA
RAAMDGLYISFGSIVTFRNAPLVQAAKYPDDRLIVETDAPFLAPVPRGRQNEPA
FVRHTAEHIAKLNQTLQVAAVTENFYRLFKKVPDMRTV"
4277..4492
/gene="NMB0772"
CDS
4277..4492
/gene="NMB0772"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF41185.1"
/db_xref="GI:7226003"
/translation="MKAIHPYACPRCRLPANTFRTCMANSAKFCIAKGRREAKKD
ESGGVALCHLPDSRIVEEWEYFSQY"
4593..4904
/gene="NMB0773"
CDS
4593..4904
/gene="NMB0773"

```

```

/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF41186.1"
/db_xref="GI:7226004"
/translation="MASTHDOIKEVWTHRVVLFMKGTQKPPQCFSSRAVOILNAAG
CTDIVVNVLENPEVRQIKEYSWPTIPQLYVNGEFVGGSDILMMEYAGELQELLK
A"
5010..5636
/gene="NMB0774"
CDS
5010..5636
/gene="NMB0774"
/notes="similar to GB:142023 SP:P43857 PID:1007128
PID:1221354 PID:1205468 percent identity: 84.13;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="uracil phosphoribosyltransferase"
/protein_id="AAF41187.1"
/db_xref="GI:7226005"
/translation="MNVNVINHPLVRHKLITLMREADCSTYKFRILATLARIMAYEAS
RDETEKYLIDWCQCIQEGDRIKGTITVPIILRAGLMDGLDLDLITAKISVVGLO
RDEETLKISYEFKFDVSDMRPALIIDPLMATGGSVATIDLLKAKGCKNKAIVLV
AAPEGVKAVNDAHDPDVTIYTAALDSHLNENGYIIPGLDAGDKIFGTR"
5659..5979
/gene="NMB0775"
CDS
5659..5979
/gene="NMB0775"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41188.1"
/db_xref="GI:7226006"
/translation="MNFQDYLATFPSIDHLGGLDVQDADGKTVHHIPAVQKGLSLKL
YNALAEFDGKLGKEAAEQGLIWFABHVAARAHFGKHPNIDLENVVGSETWLLKP
LSAQ"
6230..6649
/gene="NMB0776"
CDS
6230..6649
/gene="NMB0776"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF41189.1"
/db_xref="GI:7226007"
/translation="MKKMFLSAVLLLSAAQTWADTVFSCCKTNNKYLEVQKINRNL
VEYFSGSAKKEIALENSKADILGSRDQMGSGRWATMKFQNGEFMYTIWTGFDV
THTESGVVVERRGVEARVGCTPKTAQANFNDDFS"
6691..7431
Query Match 80.0%; Score 16.8; DB 1; Length 10579;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGCGCAAAACCTTCT 20
|||||
Db 2651 AGGAGCGCGCAAAACCATCT 2632

RESULT 11
AE009111 13051 bp DNA linear BCT 20-DEC-2001
LOCUS Agrobacterium tumefaciens strain C58 circular chromosome, section
DEFINITION 137 of 256 of the complete sequence.
ACCESSION AE009111 AE008688
VERSION AE009111.1 GI:17739946
KEYWORDS

```

SOURCE	Agrobacterium tumefaciens str. C58 (U. Washington)	gene	LIARQIGYQPNRAGVRLRTGKTNNVIALVLSDVDELMGFTSQMVFEGITEVLTSTQYHLV
ORGANISM	Agrobacterium tumefaciens str. C58 (U. Washington)		VTHPHAKDSMWPIRYILETGSDAGVLIISKIEPNDRPRVFMTEMRNPFVTHGRDMDGI
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.		EHAHPDFNDAVEAVEVLAQCRKEIAVIVPSRFSFHDHARKGNRNGRDLGLE
AUTHORS	1 (bases 1 to 13051) Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Chen, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.		MPAPK"
TITLE	The genome of the natural genetic engineer Agrobacterium tumefaciens C58		2650. .5133
JOURNAL	Science 294 (5550), 2317-2323 (2001)		/gene="manA"
MEDLINE	21608550		/note="synonym: Atul523"
PUBMED	11743193		2650. .5133
REFERENCE	2 (bases 1 to 13051) Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.		/gene="manA"
TITLE	Direct Submission		/note="identified by sequence similarity; putative; ORF
JOURNAL	Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA		located using Blastx/Glimmer"
FEATURES	98195-7242, USA		/codon_start=1
source	Location/Qualifiers		/transl_table=11
	1. .13051		/product="beta-mannosidase precursor"
	/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"		/protein_id="AAL42528.1"
	/mol_type="genomic DNA"		/db_xref="GI:17739949"
	/strain="C58"		/translation="MISSSTPETVIDLAGLHMLASVGDHATEISIPGDIIHSLKNA
	/db_xref="taxon:180835"		IIPDYHGANEKAVQWVAQQDIIERTFILDAAESWYLDIDYLDYVAIVFNDVPVL
	83. .1072		SADNCFRRYRDPIDRAVRPGENTIRIHFSNITAGAEQARQPFVYPVHGNSPIANG
	/gene="znua"		NMLRKPOCHFCDNIAIAPGLYKILKRLDTARIEHVSSQHHVPGSGEOTLYKL
	/note="synonym: Atul521"		LFAGSPASLPYLSIGDERLECGVGAGETVVRHVEVENPDLWMPAGSGLYKIL
	83. .1072		TVELPDETIVTQIGRTIELLTDKDEAGSRFAFRINGREIFCRGANWIPADALYSLT
	/gene="znua"		REKTEDLCSAVEANMMNIRVWGFGFEEDWFYDLCDRLGLLVQDWMFACNLPCSE
	/note="identified by sequence similarity; putative; ORF		DFLDNVEHYDQVRLSSHPISIALWCQDNELVGALTWFDSENNRDRILVAYDLNR
	located using Blastx/Glimmer"		TIKALKKATPEALWMPSSPASGYLDGDADHAGDSGDMHYWVSHENKSFNDYHQV
	/codon_start=1		PRFCSEFGQSYTSMPIRTVAEDKMNIAPIVIELHOKVNGNRIAGTFRYFRFP
	/transl_table=11		RDENFYLSQVQQAIAIRTAVDYRSILKPCMGLTYQLNDTWPMASVSLDYGW
	/product="ABC transporter, substrate binding protein		KALHYAARRFFQVAVSAIPADGRVTFVSMVNTAEDEVIDMKNIVALAMGNRVPLK
	[zinc]"		SAFGCTSDKAATLTDIDMSLDPGAILLAWNFIAISNGMTGEGHHVDRDYTKALELPAG
	/protein_id="AAL42526.1"		LEFSVGLNKQCFEIDVTAAGLALFIMLEADQPGRYSDNLFDLAAGETRRRIIFTPKGA
	/db_xref="GI:17739947"		GPQPHFRIFDLHTCQSSNPGETMRRKA"
	/db_xref="taxon:180835"		5237. .6268
	83. .1072		/gene="Atul524"
	/gene="znua"		5237. .6268
	/note="identified by sequence similarity; putative; ORF		/gene="Atul524"
	located using Blastx/Glimmer"		/note="identified by sequence similarity; putative; ORF
	/codon_start=1		located using Blastx/Glimmer"
	/transl_table=11		/codon_start=1
	/product="ABC transporter, substrate binding protein		/transl_table=11
	[zinc]"		/product="conserved hypothetical protein"
	/protein_id="AAL42526.1"		/protein_id="AAL42529.1"
	/db_xref="GI:17739947"		/db_xref="GI:17739950"
	/db_xref="taxon:180835"		/translation="WVQFPAENRYAMKYNHCGTKGLKLPALSLGLMNFNDTHPQT
	83. .1072		KOACIRAFDLGITHFDLANNYPGPPGSAFGEILKTFDFGYDEMIISKAGYNN
	/gene="znua"		WPGYGVGWSRKYLISCCQSLKRMGLDYDIFYSHRFPDPTPLETCGALDQIVRS
	/note="identified by sequence similarity; putative; ORF		KALYVIGSYNSKTTREAAAILKDLGTCCIIHOPYSMINRWIEDGLVDTLELIG
	located using Blastx/Glimmer"		STVFSPLAQGLMTTKYLGLGVPDGSASQSLNPAFLNERNVENIARNLSIAERRGOT
	/codon_start=1		LAQMAIAWVLRGRITISALIGASRVEQVEDCVKALDNABFSTEELAEIDRYAKDADIN
	/transl_table=11		LWAKSAERY"
	/product="ABC transporter, substrate binding protein		6460. .6786
	[zinc]"		/gene="Atul525"
	/protein_id="AAL42526.1"		6460. .6786
	/db_xref="GI:17739947"		/gene="Atul525"
	/db_xref="taxon:180835"		/note="identified by sequence similarity; putative; ORF
	83. .1072		located using Blastx/Glimmer"
	/gene="znua"		located using Blastx/Glimmer"
	/note="identified by sequence similarity; putative; ORF		/codon_start=1
	located using Blastx/Glimmer"		/transl_table=11
	/codon_start=1		/product="conserved hypothetical protein"
	/transl_table=11		/protein_id="AAL42530.1"
	/product="transcriptional regulator, LacI family"		/db_xref="GI:17739951"
	/protein_id="AAL42527.1"		/translation="MKFVTILLAAATVLAAPMAQASRHHDRHGRGVTVRQVTKKVI
	/db_xref="GI:17739948"		VKKHRWDRQRLSARERNMVDYRYYRLAEPRDQRWVRVDNQFLINAVSGLIV
	/translation="MPSAMGRPTLKTIAWTGLITVTSRALKDAPDICAETKERV		GLAAAR"
			complement (6853. .8283)
			/gene="gnd"
			/note="synonym: Atul526"
			complement (6853. .8283)
			/gene="gnd"
			/note="identified by sequence similarity; putative; ORF
			located using Blastx/Glimmer"
			/codon_start=1

```
/transl_table=11
/product="phosphogluconate dehydrogenase, decarboxylating"
/protein_id="AAL42531.1"
/db_xref="GI:17739952"
/translation="MEQAEIGLIGVWGSNLALNAEKKNIAVFNRTPEVTRKFEYA
DAGELOGOIIICETIEEFVAARPRPIIIMIKAGDPDQOQMEILKPLHANGDIMIDA
GNANFRDIIRFDNNKDSGLTFIGVSGGEGRHPSIMVGGTSDSWKRVKVLTS
ISAAYNDPCVWLYNDGNGHGVKPTIHNGIEYADNMQTAIBYGIILRDGLKMSAVEIAD
VFAENKGRNLNLYIEITEKVLRAADPITGKPMVDLIIDKAGQKGTGKWSVIEAQNMG
VAATAIEAARAILSSQKDEREAEKIPGLPTLAAAPADKKAFTADLESALLAAKVG
YAQGFATMAASKEFNWNLPMPTIIRWAGCIIRSEFLDRITSFTKDPHVANLIV
TPAFSAIVKDTDAPLRVYUAVLSGLPVSALASALGVPDARYRGRGNSALIIQAORDF
FGAHGPERDGDVKGHPGWSGADIF"
gene complement(8435..8596)
/ gene="fixS"
/ note="synonym: Atu1527"
CDS complement(8435..8596)
/ gene="fixS"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="nitrogen fixation protein FixS"
/protein_id="AAL42532.1"
/db_xref="GI:17739953"
/translation="WNMLIYLPVALFLGALGLFAPLVSRSQGQYEDMDGAARVIED
GDPRKPPS"
gene complement(8593..10884)
/ gene="fixI"
/ note="synonym: Atu1528"
CDS complement(8593..10884)
/ gene="fixI"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="nitrogen fixation protein FixI"
/protein_id="AAL42533.1"
/db_xref="GI:17739954"
/translation="MSCCAPGTEGSLGEPLNPPSSBELLASRLDQGLRQTDLSV
PCVYGACITITETALRGQVQERARVNLSSKRVAVVWKEVNGVYRTDPADIRAILA
TCYRTHLFASGQDVALRSQIIRAVALCGFASANMLLSVWSGDAADTFMPHIL
SAMIAAPALIYGRFYOASNNALXGNTNDVPILALITLSYASLNETHHGHAW
PDATVSLIFLLIGRLDHIMDKARSALAGLARISPRGATVIDAGVDYRPLADIE
FGMSIAAGRDVADVAVVCGSSDLSMIVNGESAPRVVAGDSLQAGTNLTLGSLV
AKVTASAKDSFSEVIGLMEAAEGGRARYRIADRAASYSPVHLLALVTPLGWIIF
GGDWQAMLIALAVLIITCPALGLAVPVVQVAAAGLFRRGIMVKEGSAMERLSEID
TVLPDKTGTLTGKPRLVETADVDPKPAIMAAAGLAASRHLPSLKALHAAYNGALPAYE
```

```
Query Match 80.0%; Score 16.8; DB 1; Length 13051;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGAGCGCGCAAAACCTTCTC 21
Db 7556 GCAGCGCGCAGACCTTCTC 7575
|||||
|||||

RESULT 12
BX890641
LOCUS
DEFINITION
Danio rerio clone CH211-74D4, *** SEQUENCING IN PROGRESS ***, 51
unordered pieces.
BX890641
ACCESSION
BX890641.1 GI:39923122
VERSION
HTG; HTGS PHASE1.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 173503)
McLay, K.
```

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (14-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
----- Project Information
Center project name: zc74D4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 152631 bases at least Q40
Consensus quality: 158235 bases at least Q30
Consensus quality: 161521 bases at least Q20
Insert size: 168503; sum-of-contigs
Insert size: 106745; 1.7% error; agarose-fp
Quality coverage: 2.39x in Q20 bases; sum-of-contigs Quality
coverage: 6.45x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	2982:	contig of 2982 bp in length
2983	3082:	gap of 100 bp
3083	6513:	contig of 3431 bp in length
6514	6613:	gap of 100 bp
6614	9396:	contig of 2783 bp in length
9397	9496:	gap of 100 bp
9497	12556:	contig of 3060 bp in length
12557	12656:	gap of 100 bp
12657	14890:	contig of 2234 bp in length
14891	14891:	gap of 100 bp
14891	18569:	contig of 3579 bp in length
18570	18669:	gap of 100 bp
18670	25178:	contig of 6509 bp in length
25179	25278:	gap of 100 bp
25279	28135:	contig of 2857 bp in length
28136	28235:	gap of 100 bp
28236	34813:	contig of 6578 bp in length
34814	34913:	gap of 100 bp
34914	38748:	contig of 3835 bp in length
38749	38848:	gap of 100 bp
38849	42206:	contig of 3358 bp in length
42207	42306:	gap of 100 bp
42307	48157:	contig of 5851 bp in length
48158	48257:	gap of 100 bp
48258	50455:	contig of 2198 bp in length
50456	50555:	gap of 100 bp
50556	54040:	contig of 3485 bp in length
54041	54140:	gap of 100 bp
54141	57960:	contig of 3820 bp in length
57961	58060:	gap of 100 bp
58061	63566:	contig of 5506 bp in length
63567	63666:	gap of 100 bp
63667	68425:	contig of 4759 bp in length
68426	71786:	gap of 100 bp
71787	71786:	contig of 3261 bp in length
71787	74680:	gap of 100 bp
74681	74680:	contig of 2794 bp in length
74781	82611:	contig of 7831 bp in length
82612	82711:	gap of 100 bp
82712	86166:	contig of 3455 bp in length
86167	86266:	gap of 100 bp
86267	90125:	contig of 3859 bp in length

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, X., London, P., Longacre, S., Lopez, J., Loresnuhewa, L., Louise, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Nundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Platter, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quirco, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 231028)
Worley, K.C.

Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 231028)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 14, 2002 this sequence version replaced gi:21736602.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNPC
Center clone name: CH230-77N7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 222501 bases at least Q40
Consensus quality: 224057 bases at least Q30
Consensus quality: 225007 bases at least Q20
Estimated insert size: 242141; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 224839: contig of 224839 bp in length
224840 224939: gap of unknown length
224940 226244: contig of 1305 bp in length
226245 226344: gap of unknown length
226345 227550: contig of 1206 bp in length
227551 227650: gap of unknown length
227651 231028: contig of 3378 bp in length.

FEATURES
Location/Qualifiers
1..231028
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-77N7"
1..1378
/note="wgs_contig"

misc_feature
ORIGIN
Query Match 80.0%; Score 16.8; DB 2; Length 231028;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGAGCGGCGAAACCTTCT 20
|||||
DB 8028 AGGAGCGGCGAAACCTTCT 8047
|||||

RESULT 15
BX248332
LOCUS
DEFINITION
BX248332 240116 bp DNA linear HTG 23-JUN-2003
Dario rerio clone CH211-21202, WORKING DRAFT SEQUENCE, 4 unordered pieces.
BX248332 BX248332.4 GI:30519602
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE
ORGANISM
Dario rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 240116)
McLaren, S.
Direct Submission
Submitted (20-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 10, 2003 this sequence version replaced gi:28460333.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZC21202
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 239090 bases at least Q40
Consensus quality: 239293 bases at least Q30
Consensus quality: 239474 bases at least Q20
Insert size: 239816; sum-of-contigs
Insert size: 211524; 6.0% error; agarose-fp
Quality coverage: 8.19x in Q20 bases; sum-of-contigs Quality coverage: 9.87x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 50880: contig of 50880 bp in length
 * 50881 50980: gap of 100 bp
 * 50981 65548: contig of 14568 bp in length
 * 65549 65648: gap of 100 bp
 * 65649 80411: contig of 14763 bp in length
 * 80412 80511: gap of 100 bp
 * 80512 240116: contig of 159605 bp in length.

FEATURES

source

1. .240116
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-21202"
 /clone_lib="CHORI-211"

misc_feature

1. .50880
 /note="assembly_fragment:01549
 clone_end:SP6"

vector_side:left"

misc_feature

50981..65548
 /note="assembly_fragment:01195
 fragment_chain:1"

misc_feature

65649..80411
 /note="assembly_fragment:00898
 fragment_chain:1"

misc_feature

80512..240116
 /note="assembly_fragment:00388
 fragment_chain:1
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 80.0%; Score 16.8; DB 2; Length 240116;
 Best Local Similarity 90.0%; Pred.No.1.3e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCAAAACCTTCT 20

Db 134966 AGCAGCGCGCAAAACCTTCT 134985

Search completed: May 7, 2004, 10:43:27
 Job time : 673.597 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 09:53:47 ; Search time 635.806 Seconds
(without alignments)
1363.403 Million cell updates/sec

Title: US-10-071-411A-5
Perfect score: 20
Sequence: 1 tggacttaataactttgtg 20

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	18.4	92.0	154859	9	AC012596	AC012596 Homo sapi
2	18.4	92.0	172879	2	AC107318	AC107318 Sus scrofa
3	18.4	92.0	187625	2	AC129257	AC129257 Rattus no
4	18.4	92.0	202540	2	AC022971	AC022971 Homo sapi
5	18.4	92.0	221114	2	AC022971	AC022971 Homo sapi
6	18.4	92.0	222741	2	AC128894	AC128894 Danio rer
7	18.4	92.0	226168	5	AL845284	AL845284 Zebrafish
8	18.4	92.0	251216	5	AX088593	AX088593 Zebrafish
9	18.4	92.0	263879	2	AC096060	AC096060 Rattus no
10	18.4	92.0	265893	2	AC119613	AC119613 Rattus no
11	18.4	90.0	201634	2	AC119024	AC119024 Rattus no
12	18.4	90.0	237286	2	AC113645	AC113645 Rattus no
13	18.4	90.0	239648	2	AC129771	AC129771 Rattus no
14	18.4	90.0	266099	2	AC118216	AC118216 Mus muscu
15	17.4	87.0	2108	6	AX768808	AX768808 Sequence
16	17.4	87.0	2280	10	AF057156	AF057156 Mus muscu
17	17.4	87.0	3109	1	AF269666	AF269666 Staphyloc
18	17.4	87.0	3109	6	AX144986	AX144986 Sequence
19	17.4	87.0	4216	1	AF269502	AF269502 Staphyloc
20	17.4	87.0	4216	6	AX144822	AX144822 Sequence
21	17.4	87.0	36342	9	AC093782	AC093782 Homo sapi
22	17.4	87.0	44722	9	AC107396	AC107396 Homo sapi
23	17.4	87.0	68952	2	AC101387	AC101387 Mus muscu
24	17.4	87.0	113738	2	AC138298	AC138298 Mus muscu
25	17.4	87.0	120888	2	AC119416	AC119416 Medicago
26	17.4	87.0	134571	2	BX511068	BX511068 Danio rer
27	17.4	87.0	135774	2	AC091380	AC091380 Mus muscu
28	17.4	87.0	137220	10	AL844225	AL844225 Mouse DNA
29	17.4	87.0	137237	9	AC092213	AC092213 Homo sapi
30	17.4	87.0	149015	10	AL845089	AL845089 Mus muscu
31	17.4	87.0	151923	10	AL845499	AL845499 Mouse DNA
32	17.4	87.0	156169	2	AC147095	AC147095 Pan trogl
33	17.4	87.0	157823	2	AL161795	AL161795 Homo sapi
34	17.4	87.0	162797	2	AC112974	AC112974 Mus muscu
35	17.4	87.0	165010	9	HSDJ839B4	AL109754 Human DNA
36	17.4	87.0	167361	2	BX323830	BX323830 Danio rer
37	17.4	87.0	169972	9	AL513284	AL513284 Human DNA
38	17.4	87.0	181900	10	BX284114	BX284114 Mouse DNA
39	17.4	87.0	182958	9	AC007270	AC007270 Homo sapi
40	17.4	87.0	190168	10	AC127036	AC127036 Mus muscu
41	17.4	87.0	191539	2	AC122119	AC122119 Mus muscu
42	17.4	87.0	191919	9	AC016770	AC016770 Homo sapi
43	17.4	87.0	196459	2	AC010986	AC010986 Homo sapi
44	17.4	87.0	197793	2	AC141340	AC141340 Rattus no
45	17.4	87.0	207585	2	AC123127	AC123127 Rattus no

ALIGNMENTS

RESULT 1
AC012596 AC012596 154859 bp DNA linear PRI 15-OCT-2003
LOCUS Homo sapiens BAC clone CTD-2523K17 from 7, complete sequence.
DEFINITION AC012596
ACCESSION AC012596
VERSION AC012596.4 GI:8748959
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154859)
AUTHORS Sulston, J.E. and Wilson, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

99063792
 9847074
 2 (bases 1 to 154859)
 Harkins,R., Hawkins,M., Maupin,R. and Parker,C.
 The sequence of Homo sapiens BAC clone CTD-2523K17
 Unpublished (2001)
 3 (bases 1 to 154859)
 Waterston,R.H.
 Direct Submission
 Submitted (30-OCT-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 154859)
 Waterston,R.H.
 Direct Submission
 Submitted (27-JUN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 154859)
 Waterston,R.
 Direct Submission
 Submitted (02-OCT-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 6 (bases 1 to 154859)
 Wilson,R.
 Direct Submission
 Submitted (15-OCT-2003) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jun 27, 2000 this sequence version replaced gi:7631132.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics

 Center project name: H_T02523K17

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:
 Clone CTD-2523K17 is from a release of the human BAC library CTD.
 The library contains cloned DNA from human sperm. See: Shizuya et
 al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al.,
 Genomics 34:213-8 (1996). The clone is available from Research
 Genetics, Inc. (<http://www.resgen.com>).
 VECTOR: pBelBAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 CTD-2523K17 contains a transposon in the unfinished portion of the
 clone.

The clone sequenced to the left is RP5-953B5, 200 bp overlap the
 clone sequenced to the right is GS1-512I21, 200 bp overlap. Actual
 start of this clone is at base position 96545 of RP5-953B5 actual
 end is at base position 27407 of GS1-512I21.

FEATURES

Location/Qualifiers	source
1..154859	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="CTD-2523K17"
	/clone_lib="CTD"
741..918	/rpt_family="L1"
repeat_region	
1896..2241	/rpt_family="L1"
repeat_region	
2276..2467	/rpt_family="L1"
repeat_region	
2556..2633	/rpt_family="L1"
repeat_region	
2634..2668	/rpt_family="L1"
repeat_region	
2668..2692	/rpt_family="(CA)n"
repeat_region	
2669..2845	/rpt_family="(CA)n"
repeat_region	
4383..4530	/rpt_family="L1"
repeat_region	
4789..4815	/rpt_family="(CA)n"
repeat_region	
4969..5855	/rpt_family="L1"
repeat_region	
5856..6088	/rpt_family="Alu"
repeat_region	
6155..6179	/rpt_family="AT-rich"
repeat_region	
6273..6301	/rpt_family="AT-rich"
repeat_region	
6311..6550	/rpt_family="L1"
repeat_region	
6551..7036	/rpt_family="MALR"
repeat_region	
7358..7550	/rpt_family="L1"
repeat_region	
7574..7836	/rpt_family="L1"
repeat_region	
7875..8003	/rpt_family="L1"
repeat_region	
8031..8101	/rpt_family="A-rich"
repeat_region	
8805..9052	/rpt_family="MER121"
repeat_region	
9980..10303	/rpt_family="Alu"
repeat_region	
10542..10793	/rpt_family="MIR"
repeat_region	
12536..12626	/rpt_family="MER1_type"
repeat_region	
12683..12762	/rpt_family="MER1_type"
repeat_region	
12697..12778	/rpt_family="MER1_type"
repeat_region	
12909..13069	/rpt_family="CT-rich"
repeat_region	
14153..14376	/rpt_family="MIR"
repeat_region	
14398..14645	/rpt_family="L2"
repeat_region	
14875..14954	/rpt_family="GA-rich"
repeat_region	
15352..15465	

```

/rpt_family="L2"
1515. .15757
/rpt_family="Alu"
1598. .16108
/rpt_family="L2"
16135. .16297
/rpt_family="MaLR"
16450. .16547
/rpt_family="MaLR"
16755. .16802
/rpt_family="L2"
17541. .17682
/rpt_family="Alu"
17683. .17976
/rpt_family="Alu"
17977. .18039
/rpt_family="Alu"
18339. .18622
/rpt_family="TA)n"
18743. .18796
/rpt_family="L2"
18797. .19159
/rpt_family="MaLR"
19160. .19304
/rpt_family="L2"
19311. .19782
/rpt_family="L1"
19818. .19925
/rpt_family="L1"
20081. .20207
/rpt_family="L2"
20250. .20396
/rpt_family="L2"
23314. .23370
/rpt_family="Achobo"
23517. .23646
/rpt_family="CRL"
23685. .23789
/rpt_family="MER1_type"
24418. .24762
/rpt_family="L1"

```

```

Query Match          92.0%; Score 18.4; DB 9; Length 154859;
Best local similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 TGGACTTAAATACCTTTGTG 20
Db 97872 TGGACTTATATACCTTTGTG 97891

```

```

RESULT 2
AC107318 AC107318 172879 bp DNA linear HTG 23-JAN-2003
LOCUS AC107318
DEFINITION Sus scrofa clone RP44-137G11, WORKING DRAFT SEQUENCE, 5 ordered
pieces.

```

```

ACCESSION AC107318
VERSION AC107318.2 GI:27877171
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

```

```

REFERENCE 1 (bases 1 to 172879)
AUTHORS Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Cariaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,
Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.E.,
Margulies, E.H., Masiello, C., Maskeri, B., McDowell, J.,
Paquirigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,
Reddix-Dugue, N., Schander, K., Schusler, M.G., Sison, C.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Green, E.D.
2 (bases 1 to 172879)
Direct Submission
Submitted (18-JAN-2002) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 172879)
Green, E.D.
Direct Submission
Submitted (23-JAN-2003) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
On Jan 23, 2003 this sequence version replaced gi:18201780.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: caz
Center clone name: 137G11

```

```

COMMENT

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

```

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171872 bases at least Q40
Consensus quality: 172371 bases at least Q30
Consensus quality: 172459 bases at least Q20
Insert size: 143000; agarose-fp
Insert size: 172479; sum-of-contigs
Quality coverage: 11.29x in Q20 bases; agarose-fp
Quality coverage: 9.36x in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

```

1 28086: contig of 28086 bp in length
* 28187: gap of unknown length
* 43086: contig of 14899 bp in length
* 43185: gap of unknown length
* 81263: contig of 38078 bp in length
* 81364: gap of unknown length
* 81364: contig of 17245 bp in length
* 98609: gap of unknown length
* 98709: contig of 74171 bp in length.

```

```

FEATURES
    source
        1. 172879
            /organism="Sus scrofa"
            /mol_type="genomic DNA"
            /db_xref="taxon:9823"
            /clone="RP44-137G11"
            /clone_lib="RP44"
    misc_feature
        1. 28086

```

```

/note="assembly_fragment
clone_end:SP6
vector_side:left"
28187..43085
/note="assembly_fragment"
43186..81263
/note="assembly_fragment"
81364..98608
/note="assembly_fragment"
98709..172879
/note="assembly_fragment
clone_end:T7
vector_side:right"
128354..172879
/note="clone overlaps with GenBank Accession Number
AC105368 clone RP44-249N10 (center project name cay)"

ORIGIN
Query Match      92.0%; Score 18.4; DB 2; Length 172879;
Best Local Similarity 95.0%; Pred. NO. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGACTTAATACTTTGTG 20
Db 123542 TGGACTTAATACTTTGTG 123561

RESULT 3
AC129257/c
LOCUS          187625 bp      DNA      linear      HTG 19-NOV-2002
DEFINITION    Rattus norvegicus clone CH230-340M9, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC129257
AC129257.3 GI:25072848
HTG: HTGS_PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS      Rattus norvegicus (Norway rat)
SOURCE        Rattus norvegicus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 187625)
REFERENCE
Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Aisbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,K., Garcia,A., Garner,T., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,E., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokeleneh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajk,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Staylor,T., Thomas,N., Thomas,S., Tingey,A., Tabot,P., Taylor,C.,
Taylor,K., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,J., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 187625)
Worley,K.C.
Direct Submission
Submitted (28-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 187625)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23194899.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHF
Center clone name: CH230-340M9
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 172054 bases at least Q40
Consensus quality: 174992 bases at least Q30
Consensus quality: 176501 bases at least Q20
Estimated insert size: 179977; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12674: contig of 12674 bp in length
* 12675 12774: gap of unknown length
* 12775 187625: contig of 174851 bp in length.

```

FEATURES

Location/Qualifiers
 1. .187625
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-340M9"

misc_feature

1. .1093
 /note="wgs_end_extension
 clone_end:T7"
 3339..3790
 /note="clone_boundary
 clone_end:T7"
 site:

misc_feature

end_sequence:RXAQW77TJ"
 3812..20739
 /note="clone_boundary
 clone_end:Sp6"
 site:

misc_feature

end_sequence:RXAQW77TV"
 7995..9539
 /note="wgs_end_extension
 clone_end:Sp6"
 10120..12674

misc_feature

/note="wgs_end_extension
 clone_end:Sp6"
 12775..14093

misc_feature

/note="wgs_end_extension
 clone_end:Sp6"
 182404..183798

misc_feature

/note="wgs_end_extension
 clone_end:Sp6"
 184853..187625

misc_feature

/note="wgs_end_extension
 clone_end:Sp6"

ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 187625;

Best Local Similarity 95.0%; Pred. No. 1.5e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGACTTAATACTTTTGTG 20

Db 68558 TGGAGTTAATACTTTTGTG 68539

RESULT 4

AC022971

LOCUS

AC022971 Homo sapiens chromosome 7 clone RP11-426P7 map 7, WORKING DRAFT

DEFINITION

SEQUENCE, 12 unordered pieces.

ACCESSION

AC022971.3

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 202540)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,

Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

DeArellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,

Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,

Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,

Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,

McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,

TITLE
JOURNAL

COMMENT

Norman, C.H., O'Connor, T., O'Donnell, E., Olivar, T.M., Peterson, K.,
 Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A., and Zody, M.
 Direct Submission
 Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced gi:7229884.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submission@genome.wi.mit.edu

Center project name: L6470

Center clone name: 426.p.7

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 193762 bases at least Q40

Consensus quality: 198369 bases at least Q30

Consensus quality: 199935 bases at least Q20

Insert size: 20000; agarose-fp

Quality coverage: 4.9 in Q20 bases; agarose-fp

Quality coverage: 4.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2346: contig of 2346 bp in length
 * 2347 2446: gap of 100 bp
 * 2447 7723: contig of 5277 bp in length
 * 7724 7823: gap of 100 bp
 * 7824 18624: contig of 10801 bp in length
 * 18625 18724: gap of 100 bp
 * 18725 29136: contig of 10412 bp in length
 * 29137 29236: gap of 100 bp
 * 29237 38325: contig of 9089 bp in length
 * 38326 38425: gap of 100 bp
 * 38426 47812: contig of 9387 bp in length
 * 47813 47912: gap of 100 bp
 * 47913 62442: contig of 14530 bp in length
 * 62443 62542: gap of 100 bp
 * 62543 78738: contig of 16196 bp in length
 * 78739 78838: gap of 100 bp
 * 78839 95702: contig of 16864 bp in length
 * 95703 95802: gap of 100 bp
 * 95803 119240: contig of 23438 bp in length
 * 119241 119340: gap of 100 bp
 * 119341 161336: contig of 41996 bp in length
 * 161337 161436: gap of 100 bp
 * 161437 202540: contig of 41104 bp in length.

Location/Qualifiers

1..202540

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone_lib="RP11-426P7"

/clone_lib="RP11-426P7"

/clone_lib="RP11-426P7"

/clone_lib="RP11-426P7"

/clone_lib="RP11-426P7"

/clone_lib="RP11-426P7"

FEATURES

source

```

misc_feature 1..2346
/note="assembly_fragment"
misc_feature 2447..7723
/note="assembly_fragment"
misc_feature 7824..18624
/note="assembly_fragment"
misc_feature 18725..29136
/note="assembly_fragment"
misc_feature 29237..38325
/note="assembly_fragment
clone_end:17
vector_side:right"
misc_feature 38426..47812
/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature 47913..62442
/note="assembly_fragment"
misc_feature 62543..78738
/note="assembly_fragment"
misc_feature 78839..95702
/note="assembly_fragment"
misc_feature 95803..119240
/note="assembly_fragment"
misc_feature 119341..161336
/note="assembly_fragment"
misc_feature 161437..202540
/note="assembly_fragment"

ORIGIN
Query Match 92.0%; Score 18.4; DB 2; Length 202540;
Best Local Similarity 95.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGACTTAATAACTTTTGTG 20
|||||
Db 41655 TGGACTTAATAACTTTTGTG 41674

RESULT 5
BX323070/c 221114 bp DNA linear HTG 06-NOV-2003
LOCUS
DEFINITION
Danio rerio clone DKEY-45H7, WORKING DRAFT SEQUENCE, 2 unordered
pieces.
ACCESSION BX323070
VERSION BX323070.6 GI:38201326
KEYWORDS HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 221114)
Gray E.
Direct Submission
Submitted (05-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 7, 2003 this sequence version replaced gi:38091240.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zK45H7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 220855 bases at least Q40
Consensus quality: 220908 bases at least Q30
Consensus quality: 220943 bases at least Q20
Insert size: 221014; sum-of-contigs

Insert size: 223875; 2.0% error; agarose-fp
Quality coverage: 10.0ix in Q20 bases; sum-of-contigs Quality
coverage: 9.98x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 97726: contig of 97726 bp in length
* 97727 97826: gap of 100 bp
* 97827 221114: contig of 123288 bp in length.

FEATURES
source
1..221114
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-45H7"
/clone_lib="DanioKey"
1..97726
/note="assembly_fragment:00232.0"
97827..221114
/note="assembly_fragment:00659"

misc_feature 1..97726
/note="assembly_fragment:00232.0"
misc_feature 97827..221114
/note="assembly_fragment:00659"

ORIGIN
Query Match 92.0%; Score 18.4; DB 2; Length 221114;
Best Local Similarity 95.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGACTTAATAACTTTTGTG 20
|||||
Db 110768 TGGACTTAATAACTTTTGTG 110749

RESULT 6
AC128894
LOCUS
DEFINITION
Rattus norvegicus clone CH230-188C7, *** SEQUENCING IN PROGRESS
***
ACCESSION AC128894
VERSION AC128894.3 GI:25007496
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 222741)
Muzny D, Marie J, Metzker M, Lee A, Abramson S, Adams C, Alder J,
Allen C, Allen H, Alsbrooks S, Amin A, Anguiano D,
Anyalebechi V, Aoyagi A, Ayodeji M, Baca E, Baden H,
Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F,
Biswal K, Blair J, Blankenburg K, Blyth P, Brown M,
Bryant N, Buhay C, Burch P, Burrell K, Calderon E,
Cardenas V, Carter K, Cavazos I, Ceasar H, Center A,
Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J,
Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza L,
Davila M, Davis C, Davy-Carroll L, De Anda C, Dederich D,
Delgado O, Denson S, Deramo C, Ding Y, Dinh H, Divya K,
Draper H, Dugan-Kocha S, Dunn A, Durbin K, Duval B, Eaves K,
Egan A, Escotto M, Eugene C, Evans C, Falls T, Fan G,
Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P,
Fraser C, Gabisi A, Ganta R, Garcia A, Garner T, Garza M,
Fregegeorgis E, Geer K, Gill R, Grady M, Guerra W, Guevara W,
Gunaratne P, Haaland W, Hamil C, Hamilton C, Hamilton K,
Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J,
Hernandez R, Hines S, Hladun S, Hodgson A, Hogues M,
Hollins B, Howells S, Hulyk S, Hume J, Idlebird D, Jackson A,
Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A,
Karpathy S, Kelly S, Kelly S, Khan Z, King L, Kovar C,

```

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunaagoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Fu, L.-L., Puozzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, F., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Direct Submission
Unpublished
2 (bases 1 to 222741)
Worley, K.C.

Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 222741)

REFERENCE AUTHORS TITLE JOURNAL

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 15, 2002 this sequence version replaced gi:23264875.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBAC
Center clone name: CH230-188C7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 215529 bases at least Q40
Consensus quality: 217491 bases at least Q30
Consensus quality: 218287 bases at least Q20
Estimated insert size: 224316; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 222741: contig of 222741 bp in length.

FEATURES

source

1. 222741
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-188C7"
1. 786
/note="clone boundary
clone_end:Sp6
site:
end_sequence:BH327177"
complement(215009..215469)
/note="clone boundary
clone_end:T7"
site:
end_sequence:BH327175"
216841..218543
/note="wgs end_extension
clone_end:T7"
218594..220995
/note="wgs end_extension
clone_end:T7"
221046..222741
/note="wgs end_extension
clone_end:T7"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 222741;
Best Local Similarity 95.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGACTTAAATCTTTGTG 20
DB 116226 TTGACTTAAATCTTTGTG 116245

RESULT 7

AL845284/c 226168 bp DNA linear VRT 31-OCT-2002
DEFINITION Zebrafish DNA sequence from clone DKEY-277C13, complete sequence.
ACCESSION AL845284
VERSION AL845284.4 GI:24474480
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio

REFERENCE

1 (bases 1 to 226168)
Lloyd, D.
Submitted (31-OCT-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
On Nov 1, 2002 this sequence version replaced gi:24395231.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zface@sanger.ac.uk

JOURNAL

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhixiong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rerio/fishmask.shtml DKEY-277C13 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.

```
FEATURES
    source
        1..226168
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="DKEY-277C13"
            /clone_lib="DanioKey"
```

ORIGIN

```
Query Match      92.0%; Score 18.4; DB 5; Length 226168;
Best Local Similarity 95.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGCACCTTAATACATCTTTGTG 20
    ||| ||||| ||||| ||||| |||||
Db 8979 TGCATTTAAATACATCTTTGTG 8960
```

RESULT 8

```
EX088593/c
LOCUS      251216 bp DNA linear VRT 30-NOV-2003
DEFINITION Zebrafish DNA sequence from clone DKEY-6E12 in linkage group 18, complete sequence.
```

```
ACCESSION EX088593
VERSION EX088593.11 GI:385568112
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
```

```
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 251216)
```

Brown, J.

Direct Submission

Submitted (27-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfsh-help@sanger.ac.uk

On Nov 30, 2003 this sequence version replaced gi:38198247.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfsh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhixiong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-6E12 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.

```
FEATURES
    source
        1..251216
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="DKEY-6E12"
            /clone_lib="DanioKey"
```

ORIGIN

```
Query Match      92.0%; Score 18.4; DB 5; Length 251216;
Best Local Similarity 95.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGCACCTTAATACATCTTTGTG 20
    ||| ||||| ||||| ||||| |||||
Db 8977 TGCATTTAAATACATCTTTGTG 8958
```

RESULT 9

```
AC096060/c
```

```
LOCUS      263879 bp DNA linear HTG 09-NOV-2002
DEFINITION Rattus norvegicus clone CH230-42N9, *** SEQUENCING IN PROGRESS ***, 11 unordered pieces.
```

```
ACCESSION AC096060
VERSION AC096060.6 GI:24817927
```

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 263879)

REFERENCE

AUTHORS

Muzny, D., Marie, J., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschuler, S., Amin, A., Anguitano, D., Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Louised, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, E., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Quao, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Warren, J., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE

Direct Submission

REFERENCE

Unpublished

2 (bases 1 to 263879)

AUTHORS

Worley, K.C.

TITLE

Direct Submission

JOURNAL

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 263879)

AUTHORS

Rat Genome Sequencing Consortium.

TITLE

Direct Submission

JOURNAL

Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 9, 2002 this sequence version replaced gi:22772237. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GREK

Center clone name: CH230-42N9

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 245233 bases at least Q40

Consensus quality: 247905 bases at least Q30

Consensus quality: 250127 bases at least Q20

Estimated insert size: 249743; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 244902: contig of 244902 bp in length
 * 244903 245002: gap of unknown length
 * 245003 249201: contig of 4199 bp in length
 * 249202 249301: gap of unknown length
 * 249302 250381: contig of 1080 bp in length
 * 250382 250481: gap of unknown length
 * 250482 251570: contig of 1089 bp in length
 * 251571 251670: gap of unknown length
 * 251671 252006: contig of 1236 bp in length
 * 252007 253006: gap of unknown length
 * 253007 254040: contig of 1934 bp in length
 * 254041 255040: gap of unknown length
 * 255041 256750: contig of 1710 bp in length
 * 256751 258257: contig of 1407 bp in length
 * 258258 258357: gap of unknown length
 * 258358 259662: contig of 1305 bp in length
 * 259663 259762: gap of unknown length
 * 259763 261986: contig of 2224 bp in length
 * 261987 262086: gap of unknown length
 * 262087 263879: contig of 1793 bp in length.

FEATURES

Location/Qualifiers

1..263879
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-42N9"
 1..1187
 /note="wgs contig"
 12505..14384
 /note="wgs contig"
 244179..244893
 /note="clone boundary
 clone_end:T7
 site:EcoRI
 end sequence:BH261657"
 245003..246868
 /note="wgs end extension
 clone_end:T7"

ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 263879;
 Best Local Similarity 95.0%; Pred. No. 1.4e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 TGGACTTAAATACCTTTGTG 20
 |||||
 Db 117560 TTGACTTAAATACCTTTGTG 117541

RESULT 10

AC119613/c

LOCUS

DEFINITION

Rattus norvegicus clone CH230-485D2, *** SEQUENCING IN PROGRESS

***, 2 unordered pieces.

AC119613

VERSION

AC119613.7 GI:25093224

KEYWORDS

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

linear

166583 bp

DNA

HTG 19-NOV-2002

contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GVVV
Center clone name: CH230-485D2
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 152333 bases at least Q40
Consensus quality: 154876 bases at least Q30
Consensus quality: 156588 bases at least Q20
Estimated insert size: 168008; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 164069: contig of 164069 bp in length
164070 164169: gap of unknown length
164170 166583: contig of 2414 bp in length.

FEATURES

source

Location/Qualifiers
1. .166583
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-485D2"
1. .1921
/notes="wgs_contig"
1972. .5119
/notes="wgs contig"
15775. .159433
/notes="wgs contig"
161679. .164069
/notes="wgs_contig"

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 90.0%; Score 18; DB 2; Length 166583;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GACTTAATACTTTGTG 20
|||||

Db 2018 GACTTAATACTTTGTG 2001
|||||

RESULT 11

AC119024

LOCUS

DEFINITION

AC119024

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC119024 201634 bp DNA linear HTG 20-NOV-2002
Rattus norvegicus clone CH230-431C24, *** SEQUENCING IN PROGRESS
***, 6 unordered pieces.
AC119024
AC119024.4 GI:25137781
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 201634)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,

1 (bases 1 to 166583)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,R., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerr,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwunonu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindester,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savory,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 166583)
Worley,K.C.
Direct Submission
Submitted (29-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 166583)
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:22856339.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence

REFERENCE AUTHORS

1 (bases 1 to 166583)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,R., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerr,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwunonu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindester,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savory,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE JOURNAL

2 (bases 1 to 166583)
Worley,K.C.
Direct Submission
Submitted (29-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 166583)
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:22856339.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence

REFERENCE AUTHORS

1 (bases 1 to 166583)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,R., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerr,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwunonu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindester,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savory,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

COMMENT

1 (bases 1 to 166583)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,R., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerr,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwunonu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindester,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savory,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVNK
Center clone name: CH230-431C24
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 168596 bases at least Q40
Consensus quality: 172932 bases at least Q30
Consensus quality: 175333 bases at least Q20
Estimated insert size: 176417; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 188731: contig of 188731 bp in length
* 188732 188831: gap of unknown length
* 188832 190238: contig of 1407 bp in length
* 190239 190338: gap of unknown length
* 190339 192277: contig of 1939 bp in length
* 192278 192377: gap of unknown length
* 192378 194129: contig of 1752 bp in length
* 194130 194229: gap of unknown length
* 194230 197236: contig of 3007 bp in length
* 197237 197336: gap of unknown length
* 197337 201634: contig of 4298 bp in length.
*
FEATURES             Location/Qualifiers
     source            1..201634
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-431C24"
     misc_feature      1..1187
                        /note="wgs contig"
     misc_feature      complement(181846..182731)
                        /note="clone_boundary"
                        clone_end:17
                        site:
                        end sequence:B2133306"
     misc_feature      187711..188731
                        /note="wgs end_extension"
                        clone_end:17"
ORIGIN
Query Match          90.0%; Score 18; DB 2; Length 201634;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GACTTAATAACTTTGTG 20
    |||||
Db 121338 GACTTAATAACTTTGTG 121355
    |||||
RESULT 12
AC113645/c
LOCUS      AC113645
DEFINITION Rattus norvegicus clone CH230-18E14, *** SEQUENCING IN PROGRESS
AC113645
ACCESSION AC113645
VERSION   AC113645.5 GI:25006837

```

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulsegod, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathiney, S., McLeod, M. P., McNeill, I. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemel, O., Okunolu, G., Olarinpo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajda, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 237286)
Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, W., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, F., Garza, M.,
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunatne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, D.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwakoale, O., Okwony, G., Olarnpungoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindeexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Sivatsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 237286)
Worley, K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237286)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030.
On Nov 15, 2002 this sequence version replaced gi:22856777.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GREU
Center clone name: CH230-18E14
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 229189 bases at least Q40
Consensus quality: 231234 bases at least Q30
Consensus quality: 232245 bases at least Q20
Estimated insert size: 234848; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 235611: contig of 235611 bp in length
* 235612 235711: gap of unknown length
* 235712 237286: contig of 1575 bp in length.

FEATURES
source
1..237286
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-18E14"
1..1083
/note="wgs_contig"

misc_feature
1..1083
/note="wgs_contig"

ORIGIN
Query Match 90.0%; Score 18; DB 2; Length 237286;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GACTTAAATACCTTTGTG 20
|||||
Db 162940 GACTTAAATACCTTTGTG 162923

RESULT 13
AC129771 239648 bp DNA linear HTG 09-NOV-2002
LOCUS Rattus norvegicus clone CH230-9F7, WORKING DRAFT SEQUENCE, 2
DEFINITION unordered pieces.
AC129771
AC129771.3 GI:24817954
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 239648)
Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Gantä, R., Garcia, A., Garner, T., Garza, M., Gebrat, K., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Louisedge, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Narkervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okunuwon, G., Olarinuogun, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Reeves, K., Regier, M.A., Reigh, R., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartshayn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 239648)
Worley, K.C.

Direct Submission
Submitted (02-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239648)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDEG
Center clone name: CH230-9P7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 219054 bases at least Q40
Consensus quality: 222347 bases at least Q30
Consensus quality: 224575 bases at least Q20
Estimated insert size: 226814; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 35900: contig of 35900 bp in length
* 35901 35900: gap of unknown length
* 35901 239648: contig of 203748 bp in length.
----- Location/Qualifiers
1. .239648
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-9P7"
1. .1042
/notes="wgs end extension
clone end:T7"
8344_8937
/note="clone boundary
clone end:T7"
site:EcORI
end sequence:BH341018"
24511_24545
/note="clone boundary
clone end:Sp6"
site:EcORI
end sequence:BH341020"
35901_37056
/note="wgs end extension
clone end:Sp6"
39272_41314
/note="wgs end extension
clone end:Sp6"

Query Match 90.0% Score 18; DB 2; Length 239648;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GACTTAATACTTTTGTG 20
|||||
Db 137857 GACTTAATACTTTTGTG 137874

RESULT 14
AC118216
LOCUS
DEFINITION Mas musculus chromosome 1 clone RP23-209N14 map 1, *** SEQUENCING
IN PROGRESS ***, 6 unordered pieces.
ACCESSION AC118216

```

AC118216.6 GI:37591281
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE
Mus musculus
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 266099)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 1, clone RP23-209N14
Unpublished
2 (bases 1 to 266099)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,I.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Mathews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 266099)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,I., Boukhgalter,B., Canarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Mathews,C., McCarthy,M.,
Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-OCT-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 9, 2003 this sequence version replaced gi:37515084.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23537
Center clone name: 209_N_14
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 22506: contig of 22506 bp in length
* 22507 22606: gap of 100 bp
* 22607 144177: contig of 121571 bp in length
* 144178 144277: gap of 100 bp
* 144278 164266: contig of 20149 bp in length
* 164267 164526: gap of 100 bp
* 164527 179738: contig of 15212 bp in length
* 179739 179838: gap of 100 bp
* 179839 225308: contig of 45470 bp in length
* 225309 225408: gap of 100 bp
* 225409 266099: contig of 40691 bp in length.
*
* Location/Qualifiers
* 1. 266099
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /chromosome="1"
* /map="1"
* /clone="RP23-209N14"
* /clone_lib="RPCI-23 Female Mouse BAC"

FEATURES
source
Query Match 90.0%; Score 18; DB 2; Length 266099;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
QY 3 GACTTAATACACTTTTGTG 20
Db 67351 GACTTAATACACTTTTGTG 67368

RESULT 15
AX768808/c 2108 bp DNA linear PAT 02-JUL-2003
LOCUS
Sequence 394 from Patent WO03011899.
DEFINITION
ACCESSION AX768808
VERSION AX768808.1 GI:32437083
KEYWORDS
Staphylococcus epidermidis
SOURCE
Staphylococcus epidermidis
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1 Foster,S., Mond,J., Clarke,S., McDowell,P. and Brummel,K.
AUTHORS
Antigenic polypeptides
TITLE
Patent: WO 03011899-A 394.13-FEB-2003;
JOURNAL
THE UNIVERSITY OF SHEFFIELD (GB) ; Biosynexus Incorporated (US)
FEATURES
source
1. 2108
* /organism="Staphylococcus epidermidis"
* /mol_type="unassigned DNA"
* /db_xref="taxon:1282"

ORIGIN
Query Match 87.0%; Score 17.4; DB 6; Length 2108;
Best Local Similarity 94.7%; Pred. No. 8.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACACTTTTGT 19
Db 115 TGAACCTTAAATACACTTTGT 97

Search completed: May 7, 2004, 10:43:33
JOB time : 641.806 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 08:32:52 ; Search time 142.903 Seconds

(without alignments)
594.556 Million cell updates/sec

Title: US-10-071-411a-5

Perfect score: 20

Sequence: 1 tggacttaataactttgtg 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_25Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	6	ABT11117 Human 5-1
2	18	90.0	18	6	ABT11175 Human 5-1
C 3	17.4	87.0	2108	7	ADA89733 Staphyloc
C 4	17.4	87.0	3109	4	AAS4344 S. epider
C 5	17.4	87.0	4216	4	AAS4180 S. epider
6	17	85.0	477	3	AAC39903 Arabidops
7	17	85.0	745	3	AAC52126 Arabidops
C 8	16.8	84.0	393	5	ABV61150 Human pro
9	16.8	84.0	436	7	ABX46257 Bovine ES
10	16.8	84.0	589	4	AAI16881 Probe #68
11	16.8	84.0	589	4	ABA60744 Human foe
12	16.8	84.0	589	4	AAI40635 Probe #93
13	16.8	84.0	589	4	ABA28808 Probe #72
14	16.8	84.0	589	4	AAC34919 Human bon
15	16.8	84.0	589	4	AAK09028 Human bra
16	16.8	84.0	589	4	ABX34675 Human liv
17	16.8	84.0	589	6	ABS09432 Human gen
C 18	16.8	84.0	1001	3	AAC57713 Arabidon
C 19	16.8	84.0	1001	3	AAC57714 Arabidon
C 20	16.8	84.0	1194	8	ADA32044 DNA encod
21	16.8	84.0	2000	6	ABZ16398 Arabidops
22	16.8	84.0	2092	5	ABAI5073 Human ner
23	16.8	84.0	2552	9	ADES3667 Human pro

24	16.8	84.0	2918	9	ADB61936 Human CDN
C 25	16.8	84.0	3844	5	ABV23912 Human pro
C 26	16.8	84.0	3844	5	ABV29793 Human pro
C 27	16.8	84.0	3987	9	ADC99111 Human KPP
C 28	16.8	84.0	4265	4	ABL06630 Drosophi
C 29	16.8	84.0	4855	2	AAV22252 Human WYO
C 30	16.8	84.0	10636	2	AAI13011 Enterococ
C 31	16.8	84.0	10636	6	ABS98806 Enterococ
C 32	16.8	84.0	18155	6	ADA38808 CODR4 ORF
C 33	16.8	84.0	70251	8	ADA02606 Human IL2
C 34	16.8	84.0	70251	9	ADB72344 Human IL2
C 35	16.8	84.0	91552	6	AAAD38803 BAC clone
C 36	16.8	84.0	222930	6	ABK84349 Human CDN
C 37	16.8	84.0	407	7	ABT41046 Toxicity
C 38	16.4	82.0	521	6	ABK54955 Human COL
C 39	16.4	82.0	581	5	ABV54118 Human PRO
C 40	16.4	82.0	586	4	AAH08831 Human CDN
41	16.4	82.0	50000	3	AAA96366 Polymorph
42	15.8	79.0	25	8	ACI29351 Human mic
43	15.8	79.0	364	4	AAS30845 Human CDN
44	15.8	79.0	437	4	AAI18500 Human bre
45	15.8	79.0	449	4	AAF17583 Human bre

ALIGNMENTS

RESULT 1

ABT11117
ID ABT11117 standard; DNA; 20 BP.

XX
AC ABT11117;

XX
DT 05-DEC-2002 (first entry)

XX
DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 5.

XX
KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
KW polyomitis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
ds.

XX
OS Homo sapiens.

XX
PN WO200262825-A2.

XX
PD 15-AUG-2002.

XX
PF 07-FEB-2002; 2002WO-US003546.

XX
PR 08-FEB-2001; 2001US-0267515P.

XX
PR 21-AUG-2001; 2001US-0314248P.

XX
(MILL-) MILLENNIUM PHARM INC.

XX
PI Barnes G, Meyer J;

XX
PI WPI; 2002-627522/57.

XX
PT New isolated nucleic acid molecule with an allelic variant of a

XX
PT polymorphic region of an 5-LO gene, useful for diagnosing and/or

XX
PT prognosticating disorders associated with an aberrant inflammatory

XX
PT response such as asthma.

XX
PS Claim 1; Page 234; 290pp; English.

XX
CC The invention relates to an isolated human nucleic acid molecule comprising an allelic variant of a polymorphic region of a 5-lipoxygenase (5-LO) gene, where the allelic variant comprises one or more nucleotide selected from any of 3, 20 or 21 base pair sequences, given in the specification, or their complement. The compositions and methods of the

CC present invention are useful for diagnosing and/or prognosing disorders
CC associated with an aberrant inflammatory response such as asthma,
CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
CC disease, atopic and contact dermatitis. The nucleic acid molecules can
CC also be useful for identifying an individual amongst other individuals
CC from the same species for use in forensic medicine and paternity testing.
CC This polynucleotide sequence represents DNA relating to the human 5-
XX lipoxigenase (5-LO) gene of the invention
SQ Sequence 20 BP; 5 A; 2 C; 4 G; 9 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGACTTAATACTTTGTG 20
|||||
DB 1 TGGACTTAATACTTTGTG 20
|||||

RESULT 2
ID ABT11175 standard; DNA; 18 BP.
XX ABT11175;

DT 05-DEC-2002 (first entry)

DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 65.

KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
KW ds.

XX Homo sapiens.

OS
XX WO200262825-A2.

PN
XX 15-AUG-2002.

PD
XX 07-FEB-2002; 2002WO-US003546.

PF
XX 09-FEB-2001; 2001US-0267515P.

PR
XX 21-AUG-2001; 2001US-0314248P.

PP
XX (MILL-) MILLENNIUM PHARM INC.

PT Barnes G, Meyer J;

PI WPI; 2002-627522/67.

PS New isolated nucleic acid molecule with an allelic variant of a
XX polymorphic region of an 5-LO gene, useful for diagnosing and/or
XX prognosticating disorders associated with an aberrant inflammatory
XX response such as asthma.

PS Disclosure; Page 290; 290pp; English.

CC The invention relates to an isolated human nucleic acid molecule
XX comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
XX (5-LO) gene, where the allelic variant comprises one or more nucleotide
XX selected from any of 3, 20 or 21 base pair sequences, given in the
XX specification, or their complement. The compositions and methods of the
XX present invention are useful for diagnosing and/or prognosing disorders
XX associated with an aberrant inflammatory response such as asthma,
XX bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
XX rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
XX rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory

CC disease, atopic and contact dermatitis. The nucleic acid molecules can
CC also be useful for identifying an individual amongst other individuals
CC from the same species for use in forensic medicine and paternity testing.
CC This polynucleotide sequence represents DNA relating to the human 5-
XX lipoxigenase (5-LO) gene of the invention
SQ Sequence 18 BP; 5 A; 2 C; 3 G; 8 T; 0 U; 0 Other;

Query Match 90.0%; Score 18; DB 6; Length 18;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGACTTAATACTTTTG 18
|||||
DB 1 TGGACTTAATACTTTTG 18
|||||

RESULT 3
ADA89733/c

ID ADA89733 standard; DNA; 2108 BP.

XX ADA89733;

DT 20-NOV-2003 (first entry)

DE Staphylococcus epidermidis antigenic partial gene sequence locus 6.

KW antigenic protein; vaccine; Staphylococcus epidermidis;
KW pathogenic organism; antibacterial; neuroprotective; immunosuppressive;
KW antiinflammatory; antiulcer; immunostimulant; ophthalmological;
KW pathogenic microbe; bacteraemia; septic shock; organ infection;
KW skin infection; bacterial basal colonisation; bacterial eye infection;
KW septicemia; tuberculosis; food poisoning; blood infection; peritonitis;
KW endocarditis; sepsis; meningitis; pneumonia; stomach ulcer; gonorrhoea;
KW toxic shock; necrotising fasciitis; impetigo; histoplasmosis;
KW Lyme disease; gastro-enteritis; dysentery; shigellosis; skin disorder;
KW gene; ds.

XX Staphylococcus epidermidis.

OS
XX WO2003011899-A2.

PN
XX 13-FEB-2003.

PD
XX 02-AUG-2002; 2002WO-GB003606.

PF
XX 02-AUG-2001; 2001GB-00018925.

PR
XX 09-JAN-2002; 2002GB-00000349.

PP (UYSH-) UNIV SHEPFIELD.

PT (BIOS-) BIOSYNEXUS INC.

PI Foster S, Mond J, Clarke S, McDowell P, Brummel K;

PS WPI; 2003-256434/25.

PT New antigenic polypeptides from Staphylococcus aureus or S. epidermidis,
XX useful as a vaccine for immunizing humans against e.g. bacteraemia, septic
XX shock, septicemia, tuberculosis, meningitis, pneumonia, gonorrhoea or
XX impetigo.

PS Claim 1; Page 176; 189pp; English.

CC The present invention describes an antigenic protein or its part, which
XX is for use as a vaccine. The antigenic protein is encoded by an isolated
XX DNA molecule of Staphylococcus aureus or S. epidermidis partial gene
XX sequences (designated dnaSA and dna SE, respectively; and which encodes a
XX protein expressed by a pathogenic organism. Also described: (1) a vaccine
XX composition comprising at least one antigenic protein; (2) a method of
XX immunising an animal against a disease or condition caused by a
XX pathogenic microbe by administering the antigenic protein or the vaccine;
XX (3) an antibody or its binding part obtainable by the method above; (4)
XX preparing a hybridoma cell line producing monoclonal antibodies; (5) a

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the *S. epidermidis* polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAH81454 to AAH83120, from *Staphylococcus epidermidis*. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the *S. epidermidis* polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed *S. epidermidis* genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention. N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence

PR 16-JUN-1999; 99US-0139452P.
 PR 17-JUN-1999; 99US-0139453P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 19-JUN-1999; 99US-0139455P.
 PR 20-JUN-1999; 99US-0139456P.
 PR 21-JUN-1999; 99US-0139457P.
 PR 22-JUN-1999; 99US-0139458P.
 PR 23-JUN-1999; 99US-0139459P.
 PR 24-JUN-1999; 99US-0139460P.
 PR 25-JUN-1999; 99US-0139461P.
 PR 26-JUN-1999; 99US-0139462P.
 PR 27-JUN-1999; 99US-0139463P.
 PR 28-JUN-1999; 99US-0139464P.
 PR 29-JUN-1999; 99US-0139465P.
 PR 30-JUN-1999; 99US-0139466P.
 PR 01-JUL-1999; 99US-0139467P.
 PR 02-JUL-1999; 99US-0139468P.
 PR 03-JUL-1999; 99US-0139469P.
 PR 04-JUL-1999; 99US-0139470P.
 PR 05-JUL-1999; 99US-0139471P.
 PR 06-JUL-1999; 99US-0139472P.
 PR 07-JUL-1999; 99US-0139473P.
 PR 08-JUL-1999; 99US-0139474P.
 PR 09-JUL-1999; 99US-0139475P.
 PR 10-JUL-1999; 99US-0139476P.
 PR 11-JUL-1999; 99US-0139477P.
 PR 12-JUL-1999; 99US-0139478P.
 PR 13-JUL-1999; 99US-0139479P.
 PR 14-JUL-1999; 99US-0139480P.
 PR 15-JUL-1999; 99US-0139481P.
 PR 16-JUL-1999; 99US-0139482P.
 PR 17-JUL-1999; 99US-0139483P.
 PR 18-JUL-1999; 99US-0139484P.
 PR 19-JUL-1999; 99US-0139485P.
 PR 20-JUL-1999; 99US-0139486P.
 PR 21-JUL-1999; 99US-0139487P.
 PR 22-JUL-1999; 99US-0139488P.
 PR 23-JUL-1999; 99US-0139489P.
 PR 24-JUL-1999; 99US-0139490P.
 PR 25-JUL-1999; 99US-0139491P.
 PR 26-JUL-1999; 99US-0139492P.
 PR 27-JUL-1999; 99US-0139493P.
 PR 28-JUL-1999; 99US-0139494P.
 PR 29-JUL-1999; 99US-0139495P.
 PR 30-JUL-1999; 99US-0139496P.
 PR 31-JUL-1999; 99US-0139497P.
 PR 01-AUG-1999; 99US-0139498P.
 PR 02-AUG-1999; 99US-0139499P.
 PR 03-AUG-1999; 99US-0139500P.
 PR 04-AUG-1999; 99US-0139501P.
 PR 05-AUG-1999; 99US-0139502P.
 PR 06-AUG-1999; 99US-0139503P.
 PR 07-AUG-1999; 99US-0139504P.
 PR 08-AUG-1999; 99US-0139505P.
 PR 09-AUG-1999; 99US-0139506P.
 PR 10-AUG-1999; 99US-0139507P.
 PR 11-AUG-1999; 99US-0139508P.

PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148342P.
 PR 14-AUG-1999; 99US-0148343P.
 PR 15-AUG-1999; 99US-0148344P.
 PR 16-AUG-1999; 99US-0148345P.
 PR 17-AUG-1999; 99US-0148346P.
 PR 18-AUG-1999; 99US-0148347P.
 PR 19-AUG-1999; 99US-0148348P.
 PR 20-AUG-1999; 99US-0148349P.
 PR 21-AUG-1999; 99US-0148350P.
 PR 22-AUG-1999; 99US-0148351P.
 PR 23-AUG-1999; 99US-0148352P.
 PR 24-AUG-1999; 99US-0148353P.
 PR 25-AUG-1999; 99US-0148354P.
 PR 26-AUG-1999; 99US-0148355P.
 PR 27-AUG-1999; 99US-0148356P.
 PR 28-AUG-1999; 99US-0148357P.
 PR 29-AUG-1999; 99US-0148358P.
 PR 30-AUG-1999; 99US-0148359P.
 PR 31-AUG-1999; 99US-0148360P.
 PR 01-SEP-1999; 99US-0151301P.
 PR 02-SEP-1999; 99US-0151302P.
 PR 03-SEP-1999; 99US-0151303P.
 PR 04-SEP-1999; 99US-0151304P.
 PR 05-SEP-1999; 99US-0151305P.
 PR 06-SEP-1999; 99US-0151306P.
 PR 07-SEP-1999; 99US-0151307P.
 PR 08-SEP-1999; 99US-0151308P.
 PR 09-SEP-1999; 99US-0151309P.
 PR 10-SEP-1999; 99US-0151310P.
 PR 11-SEP-1999; 99US-0151311P.
 PR 12-SEP-1999; 99US-0151312P.
 PR 13-SEP-1999; 99US-0151313P.
 PR 14-SEP-1999; 99US-0151314P.
 PR 15-SEP-1999; 99US-0151315P.
 PR 16-SEP-1999; 99US-0151316P.
 PR 17-SEP-1999; 99US-0151317P.
 PR 18-SEP-1999; 99US-0151318P.
 PR 19-SEP-1999; 99US-0151319P.
 PR 20-SEP-1999; 99US-0151320P.
 PR 21-SEP-1999; 99US-0151321P.
 PR 22-SEP-1999; 99US-0151322P.
 PR 23-SEP-1999; 99US-0151323P.
 PR 24-SEP-1999; 99US-0151324P.
 PR 25-SEP-1999; 99US-0151325P.
 PR 26-SEP-1999; 99US-0151326P.
 PR 27-SEP-1999; 99US-0151327P.
 PR 28-SEP-1999; 99US-0151328P.
 PR 29-SEP-1999; 99US-0151329P.
 PR 30-SEP-1999; 99US-0151330P.
 PR 01-OCT-1999; 99US-0157751P.
 PR 02-OCT-1999; 99US-0157752P.
 PR 03-OCT-1999; 99US-0157753P.
 PR 04-OCT-1999; 99US-0157754P.
 PR 05-OCT-1999; 99US-0157755P.
 PR 06-OCT-1999; 99US-0157756P.
 PR 07-OCT-1999; 99US-0157757P.
 PR 08-OCT-1999; 99US-0157758P.
 PR 09-OCT-1999; 99US-0157759P.
 PR 10-OCT-1999; 99US-0157760P.
 PR 11-OCT-1999; 99US-0157761P.
 PR 12-OCT-1999; 99US-0157762P.
 PR 13-OCT-1999; 99US-0157763P.
 PR 14-OCT-1999; 99US-0157764P.
 PR 15-OCT-1999; 99US-0157765P.
 PR 16-OCT-1999; 99US-0157766P.
 PR 17-OCT-1999; 99US-0157767P.
 PR 18-OCT-1999; 99US-0157768P.
 PR 19-OCT-1999; 99US-0157769P.
 PR 20-OCT-1999; 99US-0157770P.
 PR 21-OCT-1999; 99US-0157771P.
 PR 22-OCT-1999; 99US-0157772P.
 PR 23-OCT-1999; 99US-0157773P.
 PR 24-OCT-1999; 99US-0157774P.
 PR 25-OCT-1999; 99US-0157775P.
 PR 26-OCT-1999; 99US-0157776P.
 PR 27-OCT-1999; 99US-0157777P.
 PR 28-OCT-1999; 99US-0157778P.
 PR 29-OCT-1999; 99US-0157779P.
 PR 30-OCT-1999; 99US-0157780P.

Query Match 85.0%; Score 17; DB 3; Length 745;

Best Local Similarity 100.0%; Pred.No. 2.6e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ACTTAATACTTTTGTG 20

DB 695 ACTTAATACTTTTGTG 711

RESULT 8
ABV61150/c
ID ABV61150 standard; cDNA; 393 BP.
XX AC ABV61150;
XX DT 13-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 61141.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW Pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US005171.
XX PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0183862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 11613; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (i) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX Sequence 393 BP; 123 A; 87 C; 84 G; 99 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 5; Length 393;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGGACTTAAATACCTTTGTG 20
DB 366 TGGACTTAAATACCTTTGTG 347
RESULT 9
ABX46257
ID ABX46257 standard; cDNA; 436 BP.
XX AC ABX46257;
XX DT 21-FEB-2003 (first entry)
XX

DE XX Bovine EST associated with lactation/muscle/fat deposition #11422.
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX OS Bos Taurus.
XX PN US2002137139-A1.
XX PD 26-SEP-2002.
XX PF 24-SEP-2001; 2001US-00960352.
XX PR 12-JAN-1999; 99US-0115707P.
XX PR 11-JAN-2000; 2000US-00480902.
XX PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX WPI; 2003-110599/10.
XX New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and analysis,
PT cattle breeding, or for genetically improving cattle.
XX Claim 2; SEQ ID NO 11422; 245pp; English.
XX The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived from
CC cattle, and the LMFD nucleic acid can specifically hybridize to a second
CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC appearing as ABX4836-ABX49947, or complements of them. Also included are
CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC acid linked to a promoter and a 3' non- translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 15112 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid, where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMFD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the
CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC present sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docID=20020137139
XX Sequence 436 BP; 107 A; 81 C; 68 G; 180 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 7; Length 436;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGGACTTAAATACCTTTGTG 20
DB 40 TGTACATAAATACCTTTGTG 59
RESULT 10
AA116881
ID AA116881 standard; DNA; 589 BP.

```
XX AC AAI16881;
XX AC
XX PD
XX DT 12-OCT-2001 (first entry)
XX DE
XX DE Probe #6814 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD
XX PD
XX PD
XX PF 30-JAN-2001; 2001WO-US000670.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human cervical epithelial cells.
XX PS Claim 25; SEQ ID NO 6814; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENP). The present sequence is one such probe. The SENPs are derived
XX CC from human Hela cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging of
XX CC diseases of the cervix, notably cervical cancer. Note: The sequence data
XX CC for this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX CC
XX CC Sequence 589 BP; 128 A; 142 C; 186 G; 133 T; 0 U; 0 Other;
XX CC
XX CC Query Match 84.0%; Score 16.8; DB 4; Length 589;
XX CC Best Local Similarity 90.0%; Pred. No. 3.2e+02;
XX CC Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX QY 1 TGGACTTAAATACTTTGTG 20
XX DB 493 TGGACTTAAATATGTTGTG 512
XX
XX RESULT 11
XX AB60744
XX ID ABA60744 standard; DNA; 589 BP.
XX AC ABA60744;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #9049.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN
XX PD
XX PD
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human fetal liver.
XX PS Claim 1; SEQ ID NO 9049; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human fetal liver. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX CC
XX CC Sequence 589 BP; 128 A; 142 C; 186 G; 133 T; 0 U; 0 Other;
XX CC
XX CC Query Match 84.0%; Score 16.8; DB 4; Length 589;
XX CC Best Local Similarity 90.0%; Pred. No. 3.2e+02;
XX CC Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX QY 1 TGGACTTAAATACTTTGTG 20
XX DB 493 TGGACTTAAATATGTTGTG 512
XX
XX RESULT 12
XX AAI40635
XX ID AAI40635 standard; DNA; 589 BP.
XX AC AAI40635;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #9321 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD
XX PD
XX PD
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA
```

```

XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488997/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 25; SEQ ID NO 9321; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders
XX SQ Sequence 589 BP; 128 A; 142 C; 186 G; 133 T; 0 U; 0 Other;

Query Match      84.0%; Score 16.8; DB 4; Length 589;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
DB 493 TGGACTTAAATATGTTGTG 512

RESULT 13
ABA28808
ID ABA28808 standard; DNA; 589 BP.
XX AC
XX AC ABA28808;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #7274 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia,
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX FN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488999/53.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 1; SEQ ID NO 7274; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart. The
XX CC present sequence is one such probe. The probes may be used for

```

```

CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 589 BP; 128 A; 142 C; 186 G; 133 T; 0 U; 0 Other;

Query Match      84.0%; Score 16.8; DB 4; Length 589;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
DB 493 TGGACTTAAATATGTTGTG 512

RESULT 14
AAK34919
ID AAK34919 standard; DNA; 589 BP.
XX AC
XX AC AAK34919;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 9476.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX FN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488990/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO 9476; 659pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX CC the probes of the invention
XX SQ Sequence 589 BP; 128 A; 142 C; 186 G; 133 T; 0 U; 0 Other;

Query Match      84.0%; Score 16.8; DB 4; Length 589;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 TGGACTTAATAACTTTTGTG 20
Db 493 TGGACTTAATAATGTTGTG 512

RESULT 15
AAK09028
ID AAK09028 standard; DNA; 589 BP.
XX
AC AAK09028;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 9019.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00808408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
PS Example 4; SEQ ID NO 9019; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 589 BP; 128 A; 142 C; 186 G; 133 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 589;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAATAACTTTTGTG 20
Db 493 TGGACTTAATAATGTTGTG 512

Search completed: May 7, 2004, 10:10:28
Job time : 144.903 secs

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 10:10:37 ; Search time 33.5484 Seconds
(without alignments)
330.836 Million cell updates/sec

Title: US-10-071-411A-5

Perfect score: 20

Sequence: 1 tggacttaataactttgtg 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	675	4	US-09-134-000C-592
C 2	16.8	84.0	1001	4	US-09-641-638-347
C 3	16.8	84.0	1001	4	US-09-641-638-348
C 4	16.8	84.0	1194	4	US-09-328-352-3331
C 5	16.4	82.0	1456	4	US-09-976-594-779
6	15.8	79.0	449	4	US-09-222-575-68
7	15.8	79.0	449	4	US-09-389-681-68
8	15.8	79.0	449	4	US-09-620-405B-68
9	15.8	79.0	449	4	US-09-339-338-68
10	15.8	79.0	449	4	US-09-433-826B-68
11	15.8	79.0	449	4	US-09-604-287A-68
12	15.8	79.0	449	4	US-09-285-480-68
13	15.8	79.0	449	4	US-09-834-753-68
C 14	15.8	79.0	1001	4	US-09-671-317-52
C 15	15.8	79.0	4675	4	US-09-566-921-97
C 16	15.4	77.0	855	4	US-09-328-352-2433
C 17	15.4	77.0	1377	4	US-09-134-001C-1389
18	15.4	77.0	1782	2	US-08-381-603-3
19	15.4	77.0	1782	3	US-08-924-375-3
20	15.4	77.0	1782	3	US-08-685-212-3
21	15.4	77.0	1782	4	US-08-466-932A-3
22	15.4	77.0	1782	5	PCT-US94-02414-3
23	15.4	77.0	1782	5	PCT-US96-08899-3
24	15.4	77.0	2189	3	US-08-846-020A-1
25	15.4	77.0	2189	4	US-09-617-871-1
26	15.4	77.0	2356	1	US-07-821-716-3
C 27	15.4	77.0	580073	4	US-08-545-528D-1

C 28	15.4	77.0	786431	4	US-09-751-389-3	Sequence 3, Appli
C 29	15.2	76.0	3198	4	US-09-328-352-1410	Sequence 1410, Ap
C 30	15.2	76.0	6070	4	US-09-526-993-1	Sequence 1, Appli
C 31	15.2	76.0	6070	4	US-09-526-993-1	Sequence 1, Appli
C 32	15.2	76.0	90050	3	US-09-245-041-5	Sequence 5, Appli
C 33	15.2	76.0	169998	4	US-09-676-610B-24	Sequence 24, Appli
C 34	15.2	76.0	197496	4	US-09-877-177A-10	Sequence 10, Appli
C 35	15.2	76.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 36	15	75.0	400	4	US-08-956-171E-1988	Sequence 1988, Ap
C 37	15	75.0	489	4	US-09-976-594-337	Sequence 337, App
C 38	15	75.0	11970	3	US-09-345-217-1	Sequence 1, Appli
C 39	15	75.0	152331	3	US-09-128-155-16	Sequence 16, Appli
C 40	14.8	74.0	334	2	US-09-032-684-8	Sequence 8, Appli
C 41	14.8	74.0	334	4	US-09-644-460-8	Sequence 8, Appli
C 42	14.8	74.0	407	4	US-09-621-976-9413	Sequence 9413, Ap
C 43	14.8	74.0	774	4	US-09-328-352-2105	Sequence 2105, Ap
C 44	14.8	74.0	822	4	US-09-134-001C-804	Sequence 804, App
C 45	14.8	74.0	1086	4	US-09-134-000C-173	Sequence 173, App

ALIGNMENTS

RESULT 1

US-09-134-000C-592
; Sequence 592, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 592
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-592

Query Match Similarity 84.0%; Score 16.8; DB 4; Length 675;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAATAACTTTTGTG 20

Db 624 TGGACTTAAGGCTTTTGTG 643

RESULT 2

US-09-641-638-347/c
; Sequence 347, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 05/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23


```

; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 347
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 501
; OTHER INFORMATION: 12-855-194 : polymorphic base T or G
; NAME/KEY: misc binding
; LOCATION: 481..500
; OTHER INFORMATION: 12-855-194.mis1, potential
; NAME/KEY: misc binding
; LOCATION: 502..521
; OTHER INFORMATION: 12-855-194.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 674..694
; OTHER INFORMATION: upstream amplification primer, complement
; NAME/KEY: primer_bind
; LOCATION: 234..254
; OTHER INFORMATION: downstream amplification primer
; NAME/KEY: misc binding
; LOCATION: 489..513
; OTHER INFORMATION: 12-855-194 potential probe
US-09-641-638-347

Query Match      84.0%; Score 16.8; DB 4; Length 1001;
Best Local Similarity 90.0%; Pred. No. 31;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGACTTAAATAGTTTGTG 20
        |||||
Db      783 TAGACTTAATAGTTTGTG 764

RESULT 3
US-09-641-638-348/c
; Sequence 348, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 348
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 501
; OTHER INFORMATION: 12-855-288 : polymorphic base T or C
; NAME/KEY: misc binding
; LOCATION: 481..500
; OTHER INFORMATION: 12-855-288.mis1, potential

; NAME/KEY: misc binding
; LOCATION: 502..521
; OTHER INFORMATION: 12-855-288.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 768..788
; OTHER INFORMATION: upstream amplification primer, complement
; NAME/KEY: primer_bind
; LOCATION: 328..348
; OTHER INFORMATION: downstream amplification primer
; NAME/KEY: misc binding
; LOCATION: 489..513
; OTHER INFORMATION: 12-855-288 potential probe
US-09-641-638-348

Query Match      84.0%; Score 16.8; DB 4; Length 1001;
Best Local Similarity 90.0%; Pred. No. 31;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGACTTAAATAGTTTGTG 20
        |||||
Db      877 TAGACTTAAATAGTTTGTG 858

RESULT 4
US-09-328-352-3331/c
; Sequence 3331, Application US/09328352
; Patent No. 6562956
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3331
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3331

Query Match      84.0%; Score 16.8; DB 4; Length 1194;
Best Local Similarity 90.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGACTTAAATAGTTTGTG 20
        |||||
Db      831 TGGACTTAAATAGTTTGGG 812

RESULT 5
US-09-976-594-779/c
; Sequence 779, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 779
; LENGTH: 1456
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1384716.4
; NAME/KEY: unsure
```

```
; LOCATION: 1453
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-779

Query Match      82.0%; Score 16.4; DB 4; Length 1456;
Best Local Similarity 94.4%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GACTTAATAACTTTTGTG 20
Db 139 GACTTAATAATACTTTGTG 122

RESULT 6
US-09-222-575-68
; Sequence 68, Application US/09222575
; Patent No. 6387697
; GENERAL INFORMATION:
; APPLICANT: Yuqiu, Jiang
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer
; FILE REFERENCE: 210121.470
; CURRENT APPLICATION NUMBER: US/09/222,575
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Human
US-09-222-575-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAATAACTTTTGT 19
Db 349 TGGAAATTAATACTTTT 367

RESULT 7
US-09-389-681-68
; Sequence 68, Application US/09389681A
; Patent No. 6518237
; GENERAL INFORMATION:
; APPLICANT: Yuqiu, Jiang
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/389,681A
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-389-681-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAATAACTTTTGT 19
Db 349 TGGAAATTAATACTTTT 367

RESULT 8
US-09-620-405B-68
; Sequence 68, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-620-405B-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAATAACTTTTGT 19
Db 349 TGGAAATTAATACTTTT 367

RESULT 9
US-09-339-338-68
; Sequence 68, Application US/09339338A
; Patent No. 6573368
; GENERAL INFORMATION:
; APPLICANT: Yuqiu, Jiang
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE REFERENCE: 210121.470C2
; CURRENT APPLICATION NUMBER: US/09/339,338A
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-339-338-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAATAACTTTTGT 19
Db 349 TGGAAATTAATACTTTT 367

RESULT 10
US-09-433-826B-68
; Sequence 68, Application US/09433826B
; Patent No. 6579973
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
```

```
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C4
; CURRENT APPLICATION NUMBER: US/09/433,826B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 474
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-433-826B-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGACTTAATAACTTTTGT 19
DB      349 TGGATTAAATACTTTT 367

RESULT 11
US-09-604-287A-68
; Sequence 68, Application US/09604287A
; Patent No. 6586572
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuguu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-604-287A-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGACTTAATAACTTTTGT 19
DB      349 TGGATTAAATACTTTT 367

RESULT 12
US-09-285-480-68
; Sequence 68, Application US/09285480
; Patent No. 6590076
; GENERAL INFORMATION:
; APPLICANT: Yuqui, Jiang
; APPLICANT: Yuqui, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C1
; CURRENT APPLICATION NUMBER: US/09/285,480
; CURRENT FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 181
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-285-480-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGACTTAATAACTTTTGT 19
DB      349 TGGATTAAATACTTTT 367

RESULT 13
US-09-834-759-68
; Sequence 68, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuguu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-834-759-68

Query Match      79.0%; Score 15.8; DB 4; Length 449;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGACTTAATAACTTTTGT 19
DB      349 TGGATTAAATACTTTT 367

RESULT 14
US-09-671-317-52/c
; Sequence 52, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US3.CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
```

```
; SEQ ID NO 52
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 501
; OTHER INFORMATION: 10-418-177 : polymorphic base A or G
; NAME/KEY: misc_binding
; LOCATION: 481..500
; OTHER INFORMATION: 10-418-177.mis1, potential
; NAME/KEY: misc_binding
; LOCATION: 502..520
; OTHER INFORMATION: 10-418-177.mis2, complement
; NAME/KEY: primer_bind
; LOCATION: 325..342
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 659..676
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 489..513
; OTHER INFORMATION: 10-418-177 potential probe
; NAME/KEY: misc_feature
; LOCATION: 68,174,203,619,885,890,897,907,918
; OTHER INFORMATION: n=a, g, c or t
US-09-671-317-52
```

```
Query Match 79.0%; Score 15.8; DB 4; Length 1001;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 TGGACTTAAATACCTTTTGT 19
Db 951 TGGACTTAGATACATTGT 933
```

```
RESULT 15
US-09-566-921-97
; Sequence 97, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 97
; LENGTH: 4675
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6682888 234630.26
US-09-566-921-97
```

```
Query Match 79.0%; Score 15.8; DB 4; Length 4675;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 TGGACTTAAATACCTTTTGT 19
Db 3764 TGGAAATTAATACCTTTT 3782
```

Search completed: May 7, 2004, 11:56:22
Job time : 36.5484 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 11:54:38 ; Search time 140.968 Seconds
(without alignments)

642.699 Million cell updates/sec

Title: US-10-071-411A-5

Perfect score: 20

Sequence: 1 tggacttaataactttgtg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/FCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	436	9	US-09-960-352-11422
C 2	16.8	84.0	570	13	US-10-027-632-276514
C 3	16.8	84.0	570	16	US-10-027-632-276514
4	16.8	84.0	589	9	US-09-864-761-7274
5	16.8	84.0	700	13	US-10-027-632-104768
6	16.8	84.0	700	16	US-10-027-632-104768
7	16.8	84.0	729	13	US-10-027-632-11895
8	16.8	84.0	729	16	US-10-027-632-11895
9	16.8	84.0	755	13	US-10-027-632-24860
10	16.8	84.0	755	16	US-10-027-632-24860
C 11	16.8	84.0	824	13	US-10-027-632-11478
C 12	16.8	84.0	824	16	US-10-027-632-11478
13	16.8	84.0	835	13	US-10-027-632-144261
14	16.8	84.0	835	16	US-10-027-632-144261

C 15	16.8	84.0	1001	16	US-10-170-097-347	Sequence 347, App
C 16	16.8	84.0	1001	16	US-10-170-097-348	Sequence 348, App
C 17	16.8	84.0	2000	9	US-09-938-842A-4203	Sequence 4203, Ap
C 18	16.8	84.0	2000	11	US-09-938-842A-4203	Sequence 4203, Ap
C 19	16.8	84.0	2552	15	US-10-252-157-14	Sequence 14, Appl
C 20	16.8	84.0	2918	16	US-10-104-047-90	Sequence 90, Appl
C 21	16.8	84.0	3330	13	US-10-027-632-115294	Sequence 115294,
C 22	16.8	84.0	3330	16	US-10-027-632-115294	Sequence 115294,
C 23	16.8	84.0	3461	15	US-10-198-846-10971	Sequence 10971, A
C 24	16.8	84.0	3844	10	US-09-814-353-20980	Sequence 20980, A
C 25	16.8	84.0	4613	15	US-10-171-581-303	Sequence 303, App
C 26	16.8	84.0	10636	9	US-09-070-927A-74	Sequence 74, Appl
C 27	16.8	84.0	18155	13	US-10-415-058-10	Sequence 10, Appl
C 28	16.8	84.0	29945	16	US-10-085-117-58	Sequence 58, Appl
C 29	16.8	84.0	49600	17	US-10-459-262A-2	Sequence 2, Appl1
C 30	16.8	84.0	70251	12	US-10-052-482-112	Sequence 112, App
C 31	16.8	84.0	91552	13	US-10-415-058-5	Sequence 5, Appl1
C 32	16.4	82.0	407	12	US-10-152-319A-748	Sequence 748, App
C 33	16.4	82.0	521	9	US-09-919-580-425	Sequence 425, App
C 34	16.4	82.0	116585	13	US-10-087-192-133	Sequence 133, App
C 35	16	80.0	514	13	US-10-027-632-287763	Sequence 287763,
C 36	16	80.0	514	16	US-10-027-632-287763	Sequence 287763,
C 37	16	80.0	518	13	US-10-027-632-180831	Sequence 180831,
C 38	16	80.0	518	16	US-10-027-632-180831	Sequence 180831,
C 39	16	80.0	614	16	US-10-027-632-20330	Sequence 20330, A
C 40	15.8	79.0	25	15	US-10-098-263B-29342	Sequence 29342, A
C 41	15.8	79.0	317	9	US-09-796-692-6103	Sequence 6103, Ap
C 42	15.8	79.0	317	9	US-09-796-692-7131	Sequence 7131, Ap
C 43	15.8	79.0	317	15	US-10-040-862-6103	Sequence 6103, Ap
C 44	15.8	79.0	317	15	US-10-040-862-7131	Sequence 7131, Ap
C 45	15.8	79.0	317	16	US-10-057-475B-6103	Sequence 6103, Ap

ALIGNMENTS

RESULT 1
US-09-960-352-11422
; Sequence 11422, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Wyatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11422
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 49-LIB3057-025-Q1-K1-E2
US-09-960-352-11422

Query Match 84.0% Score 16.8; DB 9; Length 436;
Best Local Similarity 90.0%; Pred. No. 6e+02; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 2;

Qy 1 TGGACTTAATAACTTTTGTG 20
|||
Db 40 TGTACATAATACTTTTGTG 59
|||

RESULT 2
US-10-027-632-276514/c
; Sequence 276514, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 276514

; LENGTH: 570

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-276514

Query Match 84.0%; Score 16.8; DB 13; Length 570;

Best Local Similarity 90.0%; Pred. No. 6.2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAAATACCTTTTG 20

Db 258 TGGACTTAAATCCTTTTGAG 239

RESULT 3

US-10-027-632-276514/c

; Sequence 276514, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 276514

; LENGTH: 570

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-276514

Query Match

Best Local Similarity 84.0%; Score 16.8; DB 16; Length 570;

Best Local Similarity 90.0%; Pred. No. 6.2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAAATACCTTTTG 20

Db 258 TGGACTTAAATCCTTTTGAG 239

RESULT 4

US-09-864-761-7274

; Sequence 7274, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aecmca-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 7274

; LENGTH: 589

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AL121772.12

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.6

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.1

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.8

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1

US-09-864-761-7274

Query Match 84.0%; Score 16.8; DB 9; Length 589;
Best Local Similarity 90.0%; Pred. No. 6.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAAATACCTTTGTG 20
|||||
Db 493 TGGACTTAAATATGCTTTGTG 512

RESULT 5

US-10-027-632-104768
; Sequence 104768, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 104768
; LENGTH: 700
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-104768

Query Match 84.0%; Score 16.8; DB 13; Length 700;
Best Local Similarity 90.0%; Pred. No. 6.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAAATACCTTTGTG 20
|||||
Db 633 TGGACTTAAATGCTGTTGTG 652

RESULT 6

US-10-027-632-104768
; Sequence 104768, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 104768
; LENGTH: 700
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-104768

Query Match 84.0%; Score 16.8; DB 16; Length 700;
Best Local Similarity 90.0%; Pred. No. 6.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAAATACCTTTGTG 20
|||||
Db 633 TGGACTTAAATGCTGTTGTG 652

RESULT 7

US-10-027-632-11895
; Sequence 11895, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11895
; LENGTH: 729
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11895

Query Match 84.0%; Score 16.8; DB 13; Length 729;
Best Local Similarity 90.0%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGACTTAAATACCTTTGTG 20
|||||
Db 630 TGGACTTAAATGCTGTTGTG 649

RESULT 8

US-10-027-632-11895
; Sequence 11895, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30

;/ PRIOR APPLICATION NUMBER: US 60/218,006
;/ PRIOR FILING DATE: 2000-07-12
;/ PRIOR APPLICATION NUMBER: US 60/198,676
;/ PRIOR FILING DATE: 2000-04-20
;/ PRIOR APPLICATION NUMBER: US 60/193,483
;/ PRIOR FILING DATE: 2000-03-29
;/ PRIOR APPLICATION NUMBER: US 60/185,218
;/ PRIOR FILING DATE: 2000-02-24
;/ PRIOR APPLICATION NUMBER: US 60/167,363
;/ PRIOR FILING DATE: 1999-11-23
;/ PRIOR APPLICATION NUMBER: US 60/156,358
;/ PRIOR FILING DATE: 1999-09-28
;/ PRIOR APPLICATION NUMBER: US 60/146,002
;/ PRIOR FILING DATE: 1999-08-09
;/ NUMBER OF SEQ ID NOS: 325720
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO 11895
;/ LENGTH: 729
;/ TYPE: DNA
;/ ORGANISM: Human
US-10-027-632-411895

Query Match 84.0%; Score 16.8; DB 16; Length 729;
Best Local Similarity 90.0%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
|||||
Db 630 TGGACTTAAATGCTTTGTG 649

RESULT 9
US-10-027-632-24860
;/ Sequence 24860, Application US/10027632
;/ Publication No. US20020198371A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Wang, David G.
;/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;/ FILE REFERENCE: 108827.129
;/ CURRENT APPLICATION NUMBER: US/10/027,632
;/ CURRENT FILING DATE: 2002-04-30
;/ PRIOR APPLICATION NUMBER: US 60/218,006
;/ PRIOR FILING DATE: 2000-07-12
;/ PRIOR APPLICATION NUMBER: US 60/198,676
;/ PRIOR FILING DATE: 2000-04-20
;/ PRIOR APPLICATION NUMBER: US 60/193,483
;/ PRIOR FILING DATE: 2000-03-29
;/ PRIOR APPLICATION NUMBER: US 60/185,218
;/ PRIOR FILING DATE: 2000-02-24
;/ PRIOR APPLICATION NUMBER: US 60/167,363
;/ PRIOR FILING DATE: 1999-11-23
;/ PRIOR APPLICATION NUMBER: US 60/156,358
;/ PRIOR FILING DATE: 1999-09-28
;/ PRIOR APPLICATION NUMBER: US 60/146,002
;/ PRIOR FILING DATE: 1999-08-09
;/ NUMBER OF SEQ ID NOS: 325720
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO 24860
;/ LENGTH: 755
;/ TYPE: DNA
;/ ORGANISM: Human
US-10-027-632-24860

Query Match 84.0%; Score 16.8; DB 13; Length 755;
Best Local Similarity 90.0%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
|||||
Db 684 TGGACTCAATACATTGTG 703

RESULT 10
US-10-027-632-24860
;/ Sequence 24860, Application US/10027632
;/ Publication No. US20030204075A9
;/ GENERAL INFORMATION:
;/ APPLICANT: Wang, David G.
;/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;/ FILE REFERENCE: 108827.129
;/ CURRENT APPLICATION NUMBER: US/10/027,632
;/ CURRENT FILING DATE: 2002-04-30
;/ PRIOR APPLICATION NUMBER: US 60/218,006
;/ PRIOR FILING DATE: 2000-07-12
;/ PRIOR APPLICATION NUMBER: US 60/198,676
;/ PRIOR FILING DATE: 2000-04-20
;/ PRIOR APPLICATION NUMBER: US 60/193,483
;/ PRIOR FILING DATE: 2000-03-29
;/ PRIOR APPLICATION NUMBER: US 60/185,218
;/ PRIOR FILING DATE: 2000-02-24
;/ PRIOR APPLICATION NUMBER: US 60/167,363
;/ PRIOR FILING DATE: 1999-11-23
;/ PRIOR APPLICATION NUMBER: US 60/156,358
;/ PRIOR FILING DATE: 1999-09-28
;/ PRIOR APPLICATION NUMBER: US 60/146,002
;/ PRIOR FILING DATE: 1999-08-09
;/ NUMBER OF SEQ ID NOS: 325720
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO 24860
;/ LENGTH: 755
;/ TYPE: DNA
;/ ORGANISM: Human
US-10-027-632-24860

Query Match 84.0%; Score 16.8; DB 16; Length 755;
Best Local Similarity 90.0%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
|||||
Db 684 TGGACTCAATACATTGTG 703

RESULT 11
US-10-027-632-11478/c
;/ Sequence 11478, Application US/10027632
;/ Publication No. US20020198371A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Wang, David G.
;/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;/ FILE REFERENCE: 108827.129
;/ CURRENT APPLICATION NUMBER: US/10/027,632
;/ CURRENT FILING DATE: 2002-04-30
;/ PRIOR APPLICATION NUMBER: US 60/218,006
;/ PRIOR FILING DATE: 2000-07-12
;/ PRIOR APPLICATION NUMBER: US 60/198,676
;/ PRIOR FILING DATE: 2000-04-20
;/ PRIOR APPLICATION NUMBER: US 60/193,483
;/ PRIOR FILING DATE: 2000-03-29
;/ PRIOR APPLICATION NUMBER: US 60/185,218
;/ PRIOR FILING DATE: 2000-02-24
;/ PRIOR APPLICATION NUMBER: US 60/167,363
;/ PRIOR FILING DATE: 1999-11-23
;/ PRIOR APPLICATION NUMBER: US 60/156,358
;/ PRIOR FILING DATE: 1999-09-28
;/ PRIOR APPLICATION NUMBER: US 60/146,002
;/ PRIOR FILING DATE: 1999-08-09
;/ NUMBER OF SEQ ID NOS: 325720
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO 11478
;/ LENGTH: 824
;/ TYPE: DNA
;/ ORGANISM: Human


```
US-10-027-632-11478
Query Match      84.0%; Score 16.8; DB 13; Length 824;
Best Local Similarity 90.0%; Pred. No. 6.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
   ||||| ||||| |||||
Db 391 TGGACTTAAATGCTGTTGTG 372

RESULT 12
US-10-027-632-11478/c
; Sequence 11478, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11478
; LENGTH: 824
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11478

Query Match      84.0%; Score 16.8; DB 16; Length 824;
Best Local Similarity 90.0%; Pred. No. 6.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
   ||||| ||||| |||||
Db 391 TGGACTTAAATGCTGTTGTG 372

RESULT 13
US-10-027-632-144261
; Sequence 144261, Application US/10027632
; Publication No. US2002019871A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144261
; LENGTH: 824
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144261

Query Match      84.0%; Score 16.8; DB 16; Length 835;
Best Local Similarity 90.0%; Pred. No. 6.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
   ||||| ||||| |||||
Db 634 TGGACTTAAATGCTGTTGTG 653

RESULT 14
US-10-027-632-144261
; Sequence 144261, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144261
; LENGTH: 835
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144261

Query Match      84.0%; Score 16.8; DB 16; Length 835;
Best Local Similarity 90.0%; Pred. No. 6.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTGTG 20
   ||||| ||||| |||||
Db 634 TGGACTTAAATGCTGTTGTG 653

RESULT 15
US-10-027-632-347/c
; Sequence 347, Application US/10170097
; Publication No. US20030228582A1
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Amick
```

;; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;; FILE REFERENCE: GEN-T114XC2D1
;; CURRENT APPLICATION NUMBER: US/10/170,097
;; CURRENT FILING DATE: 2002-06-10
;; PRIOR APPLICATION NUMBER: US 09/641,638
;; PRIOR FILING DATE: 2000-08-16
;; PRIOR APPLICATION NUMBER: US 09/502,330
;; PRIOR FILING DATE: 2000-02-11
;; PRIOR APPLICATION NUMBER: US 60/133,200
;; PRIOR FILING DATE: 1999-05-07
;; PRIOR APPLICATION NUMBER: US 09/275,267
;; PRIOR FILING DATE: 1999-03-23
;; PRIOR APPLICATION NUMBER: US 60/119,917
;; PRIOR FILING DATE: 1999-02-12
;; NUMBER OF SEQ ID NOS: 1304
;; SOFTWARE: Patent.pm
;; SEQ ID NO 347
;; LENGTH: 1001
;; TYPE: DNA
;; ORGANISM: Homo Sapiens
;; FEATURE:
;; NAME/KEY: allele
;; LOCATION: 501
;; OTHER INFORMATION: 12-855-194 : polymorphic base T or G
;; FEATURE:
;; NAME/KEY: misc_binding
;; LOCATION: 481..500
;; OTHER INFORMATION: 12-855-194.misl, potential
;; FEATURE:
;; NAME/KEY: misc_binding
;; LOCATION: 502..521
;; OTHER INFORMATION: 12-855-194.mis2, potential complement
;; FEATURE:
;; NAME/KEY: primer_bind
;; LOCATION: 674..694
;; OTHER INFORMATION: upstream amplification primer, complement
;; FEATURE:
;; NAME/KEY: primer_bind
;; LOCATION: 234..254
;; OTHER INFORMATION: downstream amplification primer
;; FEATURE:
;; NAME/KEY: misc_binding
;; LOCATION: 489..513
;; OTHER INFORMATION: 12-855-194 potential probe
US-10-170-097-347

Query Match 84.0%; Score 16.8; DB 16; Length 1001;
Best Local Similarity 90.0%; Pred.No. 6.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGACTTAAATACCTTTTG 20
Db 783 TAGACTTAAATAGTTTGTG 764

Search completed: May 7, 2004, 13:42:23
Job time : 141.968 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 09:56:42 ; Search time 1370 Seconds
(without alignments)
435.944 Million cell updates/sec

Title: US-10-071-411a-5

Perfect score: 20

Sequence: 1 tggacttaataactttgtg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estom:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.4	92.0	832	29 BX171395	BX171395 Danio rer
C 2	17.4	87.0	219	10 BB148070	BB148070 BB148070
C 3	17.4	87.0	221	9 AV378631	AV378631 AV378631
C 4	17.4	87.0	250	9 AV326934	AV326934 AV326934

C	5	17.4	87.0	291	9	AA203996
C 6	17.4	87.0	295	10	BB306308	BB306308 mu28c10.i
C 7	17.4	87.0	325	10	BB120626	BB120626 BB306308
C 8	17.4	87.0	337	10	BB138348	BB138348 BB120626
C 9	17.4	87.0	338	10	BB145399	BB145399 BB138348
C 10	17.4	87.0	343	10	BB138294	BB138294 BB145399
C 11	17.4	87.0	359	13	BY159241	BY159241 BB138294
C 12	17.4	87.0	359	13	BY159275	BY159275 BY159241
C 13	17.4	87.0	361	12	BG090531	BG090531 BY159275
C 14	17.4	87.0	377	13	BY155187	BG090531 mac07c04
C 15	17.4	87.0	380	10	BB794500	BY155187 BY155187
C 16	17.4	87.0	388	13	BY525747	BB794500 BB794500
C 17	17.4	87.0	393	9	AI591508	BY525747 BY525747
C 18	17.4	87.0	412	12	BG092285	AI591508 vt19c09.x
C 19	17.4	87.0	412	13	BG436124	BG092285 mac07c04
C 20	17.4	87.0	421	13	BY154978	BY436124 BY436124
C 21	17.4	87.0	422	10	AW320682	BY154978 BY154978
C 22	17.4	87.0	429	10	BB776361	AW320682 uc22a08.y
C 23	17.4	87.0	430	12	BG348502	BB776361 BB776361
C 24	17.4	87.0	452	9	AA286167	BG348502 dg94d06.y
C 25	17.4	87.0	454	9	AI462460	AA286167 va30h01.i
C 26	17.4	87.0	469	9	AA154281	AI462460 ub73a10.x
C 27	17.4	87.0	510	10	BB795877	AA154281 mr15e06.x
C 28	17.4	87.0	516	13	EX766493	BB795877 BB795877
C 29	17.4	87.0	548	12	BM248782	EX766493 EX766493
C 30	17.4	87.0	561	28	AZ992027	BM248782 K0822C05-
C 31	17.4	87.0	623	14	CB447737	AZ992027 2M0276H02
C 32	17.4	87.0	625	14	CB447374	CB447737 701760 MA
C 33	17.4	87.0	626	28	AZ854645	CB447374 701376 MA
C 34	17.4	87.0	636	10	BB667994	AZ854645 2M0158F13
C 35	17.4	87.0	661	10	BB554730	BB667994 BB667994
C 36	17.4	87.0	661	10	BB554730	CB532087 754963 MA
C 37	17.4	87.0	688	13	CE637132	BB554730 BB554730
C 38	17.4	87.0	731	29	CE637132	BY760788 BY760788
C 39	17.4	87.0	746	29	CG932511	CE637132 tigr-gss-
C 40	17.4	87.0	766	13	EX468331	CG932511 MBEK43TR
C 41	17.4	87.0	811	13	EX622817	EX468331 EX468331
C 42	17.4	87.0	848	28	BZ566409	EX622817 EX622817
C 43	17.4	87.0	930	14	CD325303	BZ566409 pacs2-164
C 44	17.4	87.0	1008	10	BF672059	CD325303 AGENCOURT
C 45	17.4	87.0	1459	11	AK087746	BF672059 602152493
						AK087746 Mus muscu

ALIGNMENTS

RESULT 1
BX171395/c
LOCUS BX171395 832 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-172E24, genomic survey sequence.
ACCESSION BX171395
VERSION BX171395.1 GI:28003100
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
Humphray,S.J., Huckle,E. and Durham,J.L.
Direct Submission
Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Unpublished
This sequence was generated from the SP6 end of BAC 172E24. 172E24
is part of the Daniokey BAC Library created by R. Piasterk and N.V.
Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/
Location/Qualifiers
1 .832
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"

/clone="DKEY-172E24"
 /tissue type="Testis"
 /note="vector pIndigoBAC-536"

ORIGIN

Query Match 92.0%; Score 18.4; DB 29; Length 832;
 Best Local Similarity 95.0%; Pred. No. 5.3e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TGGACTTAAATACCTTTGTG 20
 |||||
 Db 567 TGGATTAAATACCTTTGTG 548
 |||||

RESULT 2

BB148070 219 bp mRNA linear EST 28-JUN-2000
 LOCUS BB148070 RIKEN full-length enriched, adult female vagina Mus
 DEFINITION musculus cDNA clone 9930105B01 3', mRNA sequence.

ACCESSION

BB148070

VERSION

BB148070.1 GI:8803007

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 219)
 Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T.,
 Carninci.P., Endo.T., Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N.,
 Hirozane.T., Hori.F., Ishii.Y., Ishikawa.J., Ishikawa.T., Itoh.M.,
 Izawa.M., Kadota.K., Kagawa.I., Kai.C., Kawai.J., Kikuchi.N.,
 Kiyosawa.H., Kojima.Y., Kondo.S., Koya.S., Kurihara.C.,
 Kusakabe.M., Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H.,
 Okazaki.Y., Ono.T., Owa.C., Saito.H., Sakai.C., Sato.K.,
 Shibata.K., Shibata.Y., Shigemoto.Y., Shingawa.A., Shiraki.T.,
 Sogabe.Y., Sugahara.Y., Suzuki.H., Suzuki.H., Tagawa.A.,
 Takahashi.F., Tomimaga.N., Toya.T., Tsunoda.Y., Watahiki.A.,
 Watanabe.S., Yamamura.T., Yamanaka.I., Yano.R., Yasunishi.A.,
 Yokota.T., Yoshida.K., Yoshiki.A., Yoshino.M., Muramatsu.M. and
 Hayashizaki.Y.
 RIKEN Mouse ESTs (Konno.H., et al.)
 Unpublished (2000)

TITLE

Unpublished (2000)

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci.P., Nishiyama.Y., Westover.A., Itoh.M., Nagaoka.S.,
 Sasaki.N., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.
 Thermostabilization and thermoactivation of thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh.M., Kitunai.T., Akiyama.J., Shibata.K., Izawa.M., Kawai.J.,
 Tomaru.Y., Carninci.P., Shibata.Y., Oza.Y., Muramatsu.M.,
 Okazaki.Y. and Hayashizaki.Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci.P. and Hayashizaki.Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES

source

1..219
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="9930105B01"
 /sex="female"

/tissue type="vagina"
 /dev stage="adult"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, adult female
 vagina"

/note="Site 1: Sall; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 185.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGATCTCGATTAAATAATATCCCCCCCCCCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 PLC I."

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 219;
 Best Local Similarity 94.7%; Pred. No. 1.6e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGACTTAAATACCTTTGTG 20

|||||

Db 22 GGAATTAATACCTTTGTG 4

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Itoh, M., Kitsuunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Orawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES	Location/Qualifiers
source	1. .221

```

1. c21
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="g130423A09"
/sex="male"
/tissue_type="cecum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male cecum"
/notes="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCGAGATCTTTTCTTTTCTTTTCTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length cDNA cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGATTATTAATATATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1. Cloning sites, 5' end: Sali; 3' end: BamHI"

```

ORIGIN

```
Query Match      87.0%; Score 17.4; DB 9; Length 221;
Best Local Similarity 94.7%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 2 GGACTTAAATACTTTTG TG 20
 ||| | | | | | | |
Dp 24 GGAATTAATACTTTTG TG 6

RESULT 4	AV326934	250 bp	mRNA	linear	EST 11-NOV-1999
LOCUS	AV326934/C				
DEFINITION	AV326934	RIKEN full-length enriched	adult	male medulla	oblongata
ACCESSION	Mus musculus	cDNA clone	G33041606	3', mRNA	sequence.
VERSION	AV326934.1				
KEYWORDS	EST	GI:6366986			

REFERENCE	EST.	SOURCE
		Mus musculus (house mouse)

ORGANISM	Mus musculus	Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	1 (bases 1 to 250)	
REFERENCE		

Authors

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Iton, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A.,

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Research	1980, Vol. 83, No. 2, pp. 11-20
3. The Importance of Parental Involvement	Journal of Educational Research	1980, Vol. 83, No. 3, pp. 21-30
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 31-40
5. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 5, pp. 41-50
6. The Impact of Socioeconomic Status on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 6, pp. 51-60
7. The Importance of Teacher Education	Journal of Educational Research	1980, Vol. 83, No. 7, pp. 61-70
8. The Effect of Teacher Attitudes on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 8, pp. 71-80
9. The Role of the School in the Development of the Child	Journal of Educational Research	1980, Vol. 83, No. 9, pp. 81-90
10. The Impact of Teacher Training on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 10, pp. 91-100

Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al. 1999)
 Unpublished (1999)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel.: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>

Sasaki, N., Izawa, M., Watanishi, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7) 3455-3460 (1998)

Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5) 463-470 (1999)

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES	SOURCE
-----------------	---------------

```
1. 250
  organism="Mus musculus"
  mol_type="rRNA"
  strain="C57BL/6J"
  db_xref="taxon:10090"
  clone="6310415D06"
  sex="male"
  tissue_type="medulla ob
  dev_stage="adult"
  lab_host="DH108"
  clone_lib="RIKEN full-l
  clone_id=""
```

000000data" 1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN, Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5',
 GAGAGAGAGAGATCCAGACGCTCTTTTCTTTTCTTCTTN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 100.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5', GAGAGAGAGATCTCGAGTTATTAATATGCCCCCCCCC
 3']. cDNA was cloned into the XhoI and BamHI sites.
 Vector: a modified pBluescript KS(+) after bulk excision
 from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
 BamHI"

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 250;
Best Local Similarity 94.7%; Pred. No. 1.6e+03;
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GGACTTAAATACCTTTGTG 20
 ||| ||||| ||||| |||||
Db 48 GGAATTAAATACCTTTGTG 30

RESULT 5					
AA203996/C					
LOCUS	AA203996	291 bp	mRNA	linear	EST 19-FEB-1997
DEFINITION	mus28c10.r1	Soares thymus	musculus	CDNA clone	IMAGE:640722

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1. .359
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I83002IM19"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow macrophage"

ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 359;
Best Local Similarity 94.7%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGACTTAATACTTTGTG 20

DB 135 GGAATTAATACTTTGTG 117

RESULT 12

BY159275/c

LOCUS BY159275 RIKEN full-length enriched, bone marrow macrophage Mus 359 bp mRNA linear EST 10-DEC-2002
musculus cDNA clone I830021019 5', mRNA sequence.

ACCESSION BY159275

VERSION BY159275.1 GI:26295921

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 359)

REFERENCE

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Otsu, K., Saito, R., Suzuki, H., Yananaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bratt, D., Brusic, V., Chothis, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gassi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sanderlin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,

Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Shuhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1. .359
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I830021019"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow macrophage"

ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 359;
Best Local Similarity 94.7%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGACTTAATACTTTGTG 20

DB 135 GGAATTAATACTTTGTG 117

```

RESULT 13
BG090531/c
LOCUS
DEFINITION
  BG090531 361 bp mRNA linear EST 26-JAN-2001
  IMAGE:3999174 5', mRNA sequence.
ACCESSION
  BG090531
VERSION
  BG090531.1 GI:12573094
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 361)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  This clone is available royalty-free through LNL; contact the
  IMAGE Consortium (info@image.lnl.gov) for further information.
  MGI:1496902
  Seq primer: -40RP from Gibco
  High quality sequence stop: 339.
  Location/Qualifiers
    1..361
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="IMAGE:3999174"
      /sex="male"
      /tissue_type="Spleen"
      /dev_stage="4 weeks"
      /lab_host="DH10B"
      /clone_lib="Soares mouse 3NBMS"
      /note="Vector: pRT3D-Pac (Pharmacia) with a modified
      polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
      was primed with a Not I - oligo(dT) primer [5'
      TGTACCAATGTCAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
      3']; double-stranded cDNA was ligated to Eco RI adaptors
      (Pharmacia), digested with Not I and cloned into the Not I
      and Eco RI sites of the modified pRT3 vector. RNA
      provided by Dr. Bertrand Jordan. Library went through
      three rounds of normalization, and was constructed by
      Bento Soares and M.Fatima Bonaldo."
ORIGIN
  Query Match 87.0%; Score 17.4; DB 12; Length 361;
  Best Local Similarity 94.7%; Pred. NO. 1.6e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGACTTAATACTTTTGTG 20
  ||| ||||| ||||| |||||
Db 243 GGAAATTAATACTTTTGTG 225

RESULT 14
BY155187/c
LOCUS
DEFINITION
  BY155187 RIKEN full-length enriched, bone marrow macrophage Mus
  musculus cDNA clone 183002H04 5', mRNA sequence.
ACCESSION
  BY155187
VERSION
  BY155187.1 GI:26291807
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 377)
  Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
  Nikaide, I., Osato, R., Saito, R., Suzuki, H., Yananaka, I.,
  Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
  Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
  Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
  Batalov, S., Beiseli, K.W., Blake, J.A., Bradt, D., Brusci, V.,
  Chotha, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
  Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
  Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
  Gustincich, S., Hirokawa, N., Jackson, J.J., Jarvis, E.D., Kanai, A.,
  Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
  Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
  Matas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
  Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
  Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
  Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
  Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
  Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
  Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
  Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
  Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
  Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
  Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
  Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
  Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
  Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
  Rogers, J., Birney, E. and Hayashizaki, Y.
  Analysis of the mouse transcriptome based on functional annotation
  of 60,720 full-length cDNAs
  Nature 420, 563-573 (2002)
  22354583
  12466851
  Contact: Yoshihide Hayashizaki
  Laboratory for Genome Exploration Research Group, RIKEN Genomic
  Sciences Center (GSC), Yokohama Institute
  The Institute of Physical and Chemical Research (RIKEN)
  1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
  Tel: 81-45-503-9222
  Fax: 81-45-503-9216
  Email: genome-res@gsc.riken.go.jp,
  URL: http://genome.gsc.riken.go.jp/
  Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
  Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
  Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
  Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
  Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
  Hayashizaki, Y. Direct Submission
  Computational Analysis of Full-Length Mouse cDNAs Compared with
  Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
  Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new
  genes. Genome Res. 10 (10), 1617-1630 (2000)
  RIKEN integrated sequence analysis (RISA) system--384-format
  sequencing pipeline with 384 multicapillary sequencer. Genome Res.
  10 (11), 1757-1771 (2000)
  Computer-based methods for the mouse full-length cDNA
  encyclopedia: real-time sequence clustering for construction of a
  nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
  cDNA library was prepared and sequenced in Mouse Genome
  Encyclopedia Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in RIKEN.
  Division of Experimental Animal Research in Riken contributed to
  prepare mouse tissues.
  Tissues were provided by David A. Hume ( Depts. of Biochemistry
  and Microbiology/Parasitology Institute for Molecular Bioscience
  University of Queensland Brisbane, Q 4072 Australia ) whose
  assistance we gratefully acknowledge.
  Please visit our web site (http://genome.gsc.riken.go.jp) for
  further details.
  Location/Qualifiers
    1..377
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="183002H04"
  FEATURES
    source

```

```

/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow
macrophage"

ORIGIN
Query Match      87.0%; Score 17.4; DB 13; Length 377;
Best Local Similarity 94.7%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGACTTAATACTTTGTG 20
    ||| ||||| ||||| |||||
Db 140 GGAATTAATACTTTGTG 122

RESULT 15
BB794500/c
LOCUS BB794500
DEFINITION BB794500 RIKEN full-length enriched, embryo RCB-0549 Cle-H3 cDNA
ACCESSION BB794500
VERSION BB794500.1 GI:16964119
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 380)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tonaru,A., Toyota,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
TITLE Unpublished (2001)
JOURNAL Contact: Yoshihide Hayashizaki
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
Genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1. .380
/organism="Mus musculus"
/mol_type="mRNA"

FEATURES
source

```

```

/db_xref="taxon:10090"
/clone="G430150K05"
/cell_line="RCB-0549 Cle-H3"
/dev_stage="embryo"
/clone_lib="RIKEN full-length enriched, embryo RCB-0549
Cle-H3 cDNA"

ORIGIN
Query Match      87.0%; Score 17.4; DB 10; Length 380;
Best Local Similarity 94.7%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGACTTAATACTTTGTG 20
    ||| ||||| ||||| |||||
Db 185 GGAATTAATACTTTGTG 167

Search completed: May 7, 2004, 11:54:25
Job time : 1374 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 7, 2004, 09:53:47 ; Search time 667.597 Seconds
(without alignments)
1363.403 Million cell updates/sec

Title: US-10-071-411a-6

Perfect score: 21

Sequence: 1 tcattgtatccaattagagact 21

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_atg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vt.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vt.*

29: em_vt.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_man.*

37: em_htg_vrt.*

38: em_ey.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	19.4	92.4	2189	6	AR199380	AR199380 Sequence
2	19.4	92.4	2189	9	AX285281	AX285281 Sequence
3	19.4	92.4	2189	9	HUMLIP05	M38191 Human 5-lip
4	19.4	92.4	129266	9	AL731567	AL731567 Human DNA
5	19.4	92.4	160654	2	AC011879	AC011879 Homo sapi
6	18.4	87.6	190223	9	AL355137	AL355137 Human DNA
7	18.4	87.6	219241	10	AC116458	AC116458 Mus muscu
8	17.8	84.8	22996	9	AC087175	AC087175 Homo sapi
9	17.8	84.8	85598	9	AC086584	AC086584 Homo sapi
10	17.8	84.8	85624	9	AL133933	AL133933 Human DNA
11	17.8	84.8	95391	9	AC005926	AC005926 Homo sapi
12	17.8	84.8	148509	9	AC068014	AC068014 Homo sapi
13	17.8	84.8	160256	10	AC122451	AC122451 Mus muscu
14	17.8	84.8	172657	9	AC004386	AC004386 Homo sapi
15	17.8	84.8	176159	10	AL772220	AL772220 Mouse DNA
16	17.8	84.8	178414	2	AC137801	AC137801 Homo sapi
17	17.8	84.8	184689	2	AC009172	AC009172 Homo sapi
18	17.8	84.8	186134	9	AC007598	AC007598 Homo sapi
19	17.8	84.8	186559	2	AC137808	AC137808 Homo sapi
20	17.8	84.8	187098	2	AC015984	AC015984 Homo sapi
21	17.8	84.8	191212	2	AC129856	AC129856 Rattus no
22	17.8	84.8	196646	2	AC073146	AC073146 Homo sapi
23	17.8	84.8	197140	9	AC145966	AC145966 Pan trogl
24	17.8	84.8	200279	2	AC109165	AC109165 Mus muscu
25	17.8	84.8	207345	9	AC009062	AC009062 Homo sapi
26	17.8	84.8	212835	2	AC007612	AC007612 Homo sapi
27	17.8	84.8	215435	10	AL662838	AL662838 Mouse DNA
28	17.8	84.8	234885	2	AC097033	AC097033 Rattus no
29	17.8	84.8	239661	2	AC132157	AC132157 Rattus no
30	17.8	84.8	252509	2	AC121441	AC121441 Rattus no
31	17.8	84.8	254019	2	AC120470	AC120470 Rattus no
32	17.8	84.8	258328	2	AC133789	AC133789 Rattus no
33	17.8	84.8	260781	2	AC098031	AC098031 Rattus no
34	17.4	82.9	290	11	BV045235	BV045235 S212P6339
35	17.4	82.9	70109	2	AC101413	AC101413 Mus muscu
36	17.4	82.9	95672	10	AL844604	AL844604 Mouse DNA
37	17.4	82.9	150816	2	AC119532	AC119532 Rattus no
38	17.4	82.9	167733	9	AC008270	AC008270 Homo sapi
39	17.4	82.9	170336	10	AC125137	AC125137 Mus muscu
40	17.4	82.9	172947	2	AC102325	AC102325 Mus muscu
41	17.4	82.9	178679	2	AC113946	AC113946 Mus muscu
42	17.4	82.9	201451	2	AC114041	AC114041 Rattus no
43	17.4	82.9	202162	10	AL845318	AL845318 Mouse DNA
44	17.4	82.9	214488	10	AL773520	AL773520 Mouse DNA
45	17.4	82.9	226950	10	AL671011	AL671011 Mouse DNA

ALIGNMENTS

RESULT 1
AR199380
LOCUS AR199380 2189 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355434.
ACCESSION AR199380
VERSION AR199380.1 GI:20249454
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2189)
AUTHORS Drazen, J. M., In, K.-H., Asano, K., Beier, D. and Grobholz, J.
TITLE 5-Lipoxygenase gene polymorphisms and their use in classifying Patients
JOURNAL Patent: US 6355434-A 1 12-MAR-2002;

```

FEATURES
    source          Location/Qualifiers
    1..2189
    /organism="unknown"
    /mol_type="unassigned DNA"

ORIGIN
Query Match          92.4%; Score 19.4; DB 6; Length 2189;
Best Local Similarity 95.2%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
Db 549 TCATGTATCCGATTAGAGACT 569

RESULT 2
AX285281          2189 bp          DNA          linear          PAT 20-NOV-2001
LOCUS             Sequence 5 from Patent WO0179560.
DEFINITION        AX285281
ACCESSION         AX285281
VERSION           AX285281.1 GI:17045963
KEYWORDS
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS           Anderson, W.H., Edwards, L.D., Emmett, A.H., Pillai, S. and
                  Sprankel, C.S.
TITLE             Medicine response assay in respiratory disease
JOURNAL           Patent: WO 0179560-A 5 25-OCT-2001;
                  GLAXO GROUP LIMITED (GB)
FEATURES
    source          Location/Qualifiers
    1..2189
    /organism="Homo sapiens"
    /mol_type="unassigned DNA"
    /db_xref="taxon:9606"

ORIGIN
Query Match          92.4%; Score 19.4; DB 6; Length 2189;
Best Local Similarity 95.2%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
Db 549 TCATGTATCCGATTAGAGACT 569

RESULT 3
HUMLIPO5
LOCUS             Human 5-lipoxygenase gene, exon 1.
DEFINITION        M38191
ACCESSION         M38191
VERSION           M38191.1 GI:187166
KEYWORDS          5-lipoxygenase.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS           Hoshiko, S., Radmark, O. and Samuelsson, B.
TITLE             Characterization of the human 5-lipoxygenase gene promoter
JOURNAL           Proc. Natl. Acad. Sci. U.S.A. 87 (23), 9073-9077 (1990)
MEDLINE           91067649
PUBMED            2251250
COMMENT           Original source text: Human DNA, clone lx12A.
                  Draft entry and computer-readable sequence for [Proc. Natl. Acad.
                  Sci. U.S.A. (1990) in press] kindly submitted
                  by O.Radmark, 28-AUG-1990.
FEATURES
    source          Location/Qualifiers
    1..2189
    /organism="Homo sapiens"
    /mol_type="genomic DNA"

FEATURES
    source          Location/Qualifiers
    1..2189
    /organism="unknown"
    /mol_type="unassigned DNA"

ORIGIN
Query Match          92.4%; Score 19.4; DB 9; Length 2189;
Best Local Similarity 95.2%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCATGTATCCCAATTAGAGACT 21
Db 549 TCATGTATCCGATTAGAGACT 569

RESULT 4
AL731567          129266 bp        DNA          linear          PRI 20-JUN-2002
LOCUS             Human DNA sequence from clone Rp11-67C2 on chromosome 10, complete
DEFINITION        AL731567 AC010865
ACCESSION         AL731567
VERSION           AL731567.6 GI:21537524
KEYWORDS          HTG.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS           Whitehead, S.
TITLE             Direct Submission
JOURNAL           Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
                  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                  humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
                  On Jun 21, 2002 this sequence version replaced gi:21213582.
                  Draft Sequence produced by Genome Therapeutics Corp, 100 Beaver
                  Street, Waltham, MA 02453, USA
                  http://www.genomecorp.com
COMMENT           During sequence assembly data is compared from overlapping clones.
                  Where differences are found these are annotated as variations
                  together with a note of the overlapping clone name. Note that the
                  variation annotation may not be found in the sequence submission
                  corresponding to the overlapping clone, as we submit sequences with
                  only a small overlap as described above.
                  This sequence was finished as follows unless otherwise noted: all
                  regions were either double-stranded or sequenced with an alternate
                  chemistry or covered by high quality data (i.e., phred quality >=
                  30); an attempt was made to resolve all sequencing problems, such
                  as compressions and repeats; all regions were covered by at least
                  one plasmid subclone or more than one M13 subclone; and the
                  assembly was confirmed by restriction digest. The following
                  abbreviations are used to associate primary accession numbers given
                  in the feature table with their source databases: EMBL, SW,
                  SWISSPROT, Tr, TREMBL, Wp, WORMPEP; Information on the WORMPEP
                  database can be found at
                  http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
                  was generated from part of bacterial clone contigs of human
                  chromosome 10, constructed by the Sanger Centre Chromosome 10
                  Mapping Group. Further information can be found at

```

<http://www.sanger.ac.uk/HGP/Chr10>
 RP11-67C2 is from the library RPC1-11.1 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

FEATURES

Location/Qualifiers
 1..129266
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone=RP11-67C2
 /clone_lib="RPC1-11.1"

ORIGIN

Query Match 92.4%; Score 19.4; DB 9; Length 129266;
 Best Local Similarity 95.2%; Pred. No. 16;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 TCATGTATCCCAATTAGAGACT 21
 Db 33286 TCATGTATCCGATTAGAGACT 33306

RESULT 5

AC011879 160654 bp DNA linear HTG 16-MAR-2000
 LOCUS
 DEFINITION Homo sapiens clone RP11-16P14, WORKING DRAFT SEQUENCE, 30 unordered
 pieces.

ACCESSION

AC011879

VERSION

AC011879.3 GI:7239554

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 160654)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Becker, R., Boguski, M., Collins, S., Collins, S.,
 Cooke, P., Dearlano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
 Ferreira, P., FitzHugh, M., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gargyala, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lechoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 14, 2000 this sequence version replaced gi:6524208.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: 16P14

Center clone name: 16P14

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
 Consensus quality: 111055 bases at least Q40
 Consensus quality: 135066 bases at least Q30
 Consensus quality: 147921 bases at least Q20
 Insert size: 163000; agarose-fp
 Insert size: 157754; sum-of-contigs
 Quality coverage: 2.9 in Q20 bases; agarose-fp
 Quality coverage: 3.0 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 30 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 151: contig of 151 bp in length
 * 152: gap of 100 bp
 * 252: contig of 1509 bp in length
 * 1761: gap of 100 bp
 * 1861: contig of 1209 bp in length
 * 3070: gap of 100 bp
 * 3170: contig of 1551 bp in length
 * 4721: gap of 100 bp
 * 4821: contig of 1354 bp in length
 * 6175: gap of 100 bp
 * 6275: contig of 1143 bp in length
 * 7418: gap of 100 bp
 * 7518: contig of 1641 bp in length
 * 9159: gap of 100 bp
 * 9259: contig of 1607 bp in length
 * 10866: gap of 100 bp
 * 10966: contig of 1894 bp in length
 * 12860: gap of 100 bp
 * 12960: contig of 2712 bp in length
 * 15672: gap of 100 bp
 * 15772: contig of 2311 bp in length
 * 18083: gap of 100 bp
 * 18183: contig of 2341 bp in length
 * 20524: gap of 100 bp
 * 20624: contig of 2280 bp in length
 * 22904: gap of 100 bp
 * 23004: contig of 668 bp in length
 * 23671: gap of 100 bp
 * 23772: contig of 1770 bp in length
 * 25542: gap of 100 bp
 * 25642: contig of 2682 bp in length
 * 28324: gap of 100 bp
 * 28424: contig of 3075 bp in length
 * 31498: gap of 100 bp
 * 31599: contig of 5028 bp in length
 * 36627: gap of 100 bp
 * 36727: contig of 5383 bp in length
 * 42110: gap of 100 bp
 * 42210: contig of 6130 bp in length
 * 48340: gap of 100 bp
 * 48440: contig of 6894 bp in length
 * 55334: gap of 100 bp
 * 55434: contig of 8161 bp in length
 * 63595: gap of 100 bp
 * 63695: contig of 10249 bp in length
 * 73943: gap of 100 bp
 * 73944: contig of 9622 bp in length
 * 74044: gap of 100 bp
 * 83666: contig of 11557 bp in length
 * 83766: gap of 100 bp
 * 95423: contig of 12981 bp in length
 * 108404: gap of 100 bp
 * 108504: contig of 11717 bp in length
 * 120221: gap of 100 bp
 * 120321: contig of 12638 bp in length
 * 132959: gap of 100 bp
 * 133058: contig of 100 bp

* 133059 145697: contig of 12639 bp in length
* 145698 145797: gap of 100 bp
* 145798 160654: contig of 14857 bp in length.

FEATURES

source
1..160654
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-16P14"
/clone_lib="RPC1-11 Human Male BAC"
1..151
/note="assembly_fragment
clone end:T7
vector side:right"
252..1760
/note="assembly_fragment"
1861..3069
/note="assembly_fragment"
3170..4720
/note="assembly_fragment"
4821..6174
/note="assembly_fragment"
6275..7417
/note="assembly_fragment"
7518..9158
/note="assembly_fragment"
9259..10865
/note="assembly_fragment"
10966..12859
/note="assembly_fragment"
12960..15671
/note="assembly_fragment"
15772..18082
/note="assembly_fragment"
18183..20523
/note="assembly_fragment"
20624..22903
/note="assembly_fragment"
23004..23671
/note="assembly_fragment
clone end:SP6
vector side:right"
23772..25541
/note="assembly_fragment"
25642..28323
/note="assembly_fragment"
28424..31498
/note="assembly_fragment"
31599..36626
/note="assembly_fragment"
36727..42109
/note="assembly_fragment"
42210..48339
/note="assembly_fragment"
48440..55333
/note="assembly_fragment"
55434..63594
/note="assembly_fragment"
63695..73943
/note="assembly_fragment"
74044..83665
/note="assembly_fragment"
83766..95322
/note="assembly_fragment"
95423..108403
/note="assembly_fragment"
108504..120220
/note="assembly_fragment"
120321..132958
/note="assembly_fragment"
133059..145697
/note="assembly_fragment"
145798..160654

ORIGIN

Query Match 92.4%; Score 19.4; DB 2; Length 160654;
Best Local Similarity 95.2%; Pred.No.16;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCATGTATCCAAATTAGAGACT 21
|||||
Db 40363 TCATGTATCCGATTAGAGACT 40383
|||||

RESULT 6

AL355137 190223 bp DNA linear PRI 29-MAR-2001
LOCUS Human DNA sequence from clone RP11-511D14 on chromosome 6p23-24.3,
DEFINITION complete sequence.

ACCESSION AL355137

VERSION AL355137.23 GI:13446430

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 190223)

AUTHORS Tracey, A.

TITLE Direct Submission

JOURNAL Submitted (28-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

COMMENT On Mar 25, 2001 this sequence version replaced gi:13398746.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em: EMBL; Sw:

SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/Projects/c_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

RP11-511D14 is from the library RPC1-11.2 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-511D14 The true

left end of clone RP11-124B11 is at 63342 in this sequence. The

true right end of clone RP11-97A19 is at 130850 in this sequence.

FEATURES

source

1..190223

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="6"

/map="p23-24.3"

/clone="RP11-511D14"

/clone_lib="RPC1-11.2"

repeat_region 1..127

/note="MER5B repeat: matches 25..149 of consensus"

repeat_region 318..458

/note="LINE2 repeat: matches 6005..6156 of consensus"


```
repeat_region 904. .947
/note="L11 copies 4 mer catt 81% conserved"
repeat_region 1228. .1331
/note="L2 repeat: matches 2178. .2300 of consensus"
repeat_region 1332. .1409
/note="39 copies 2 mer tt 64% conserved"
repeat_region 4136. .4418
/note="AluJb repeat: matches 1. .298 of consensus"
repeat_region 4456. .4581
/note="MIR repeat: matches 119. .256 of consensus"
repeat_region 4837. .4955
/note="FIAM_C repeat: matches 1. .118 of consensus"
repeat_region 5057. .5134
/note="L2 repeat: matches 2106. .2187 of consensus"
repeat_region 5253. .5388
/note="MER5A repeat: matches 2. .136 of consensus"
repeat_region 5684. .6208
/note="MER6A repeat: matches 43. .605 of consensus"
repeat_region 6320. .6631
/note="AluY repeat: matches 1. .309 of consensus"
repeat_region 6798. .7088
/note="AluX repeat: matches 1. .296 of consensus"
repeat_region 7089. .7110
/note="L11 copies 2 mer ga 100% conserved"
repeat_region 7278. .7555
/note="L39 copies 2 mer tt 65% conserved"
repeat_region 7378. .7557
/note="45 copies 4 mer ttcc 84% conserved"
repeat_region 7561. .7877
/note="AluJo repeat: matches 1. .308 of consensus"
repeat_region 7931. .8145
/note="L2 repeat: matches 2483. .2709 of consensus"
repeat_region 9573. .9775
/note="LIP14 repeat: matches 5951. .6149 of consensus"
repeat_region 9777. .10160
/note="MSTA repeat: matches 1. .426 of consensus"
repeat_region 10161. .11799
/note="MSTA-internal repeat: matches 1. .1651 of consensus"
repeat_region 11800. .12189
/note="MSTA repeat: matches 1. .426 of consensus"
repeat_region 12190. .12825
/note="LIP14 repeat: matches 5318. .5954 of consensus"
repeat_region 12834. .13132
/note="AluX repeat: matches 1. .307 of consensus"
repeat_region 14975. .15059
/note="L2 repeat: matches 2049. .2140 of consensus"
repeat_region 15645. .15711
/note="L2 repeat: matches 2593. .2695 of consensus"
repeat_region 16008. .16078
/note="LIME3 repeat: matches 6029. .6100 of consensus"
repeat_region 16106. .16186
/note="FIAM repeat: matches 58. .103 of consensus"
repeat_region 16127. .16194
/note="Sequence from uni-directional dGTP reads only."
repeat_region 16160. .16225
/note="FIAM_A repeat: matches 55. .120 of consensus"
repeat_region 16504. .16657
/note="MER5B repeat: matches 20. .167 of consensus"
repeat_region 17098. .17136
/note="AluJo repeat: matches 132. .170 of consensus"
repeat_region 17137. .17444
/note="AluX repeat: matches 1. .308 of consensus"
repeat_region 17445. .17577
/note="AluJo repeat: matches 1. .132 of consensus"
repeat_region 18837. .19006
/note="L2 repeat: matches 1394. .1550 of consensus"
repeat_region 19007. .19304
/note="AluJo repeat: matches 1. .295 of consensus"
repeat_region 19305. .19621
/note="L2 repeat: matches 1550. .1918 of consensus"
repeat_region 19950. .20032
/note="L2 repeat: matches 2666. .2750 of consensus"
repeat_region 20285. .20645
/note="L2 repeat: matches 18. .411 of consensus"
20698. .20889
/note="MER58A repeat: matches 2. .223 of consensus"
21111. .21830
/note="L2 repeat: matches 1997. .2750 of consensus"
21886. .22080
/note="MER30 repeat: matches 1. .213 of consensus"
22081. .22350
/note="L2 repeat: matches 1657. .1936 of consensus"
22608. .22914
/note="AluSg repeat: matches 1. .308 of consensus"
23765. .23812
/note="L2 copies 4 mer ttgt 95% conserved"
23867. .24244
/note="MLT1B repeat: matches 1. .384 of consensus"
24245. .24280
/note="9 copies 4 mer gtgt 83% conserved"
24306. .24355
/note="25 copies 2 mer ac 74% conserved"
25201. .25236
/note="18 copies 2 mer gt 100% conserved"
25238. .25518
/note="AluJb repeat: matches 6. .286 of consensus"
25587. .25716
/note="MER5A repeat: matches 26. .178 of consensus"
25938. .26114
/note="MIR repeat: matches 80. .257 of consensus"
26276. .26340
/note="L2 repeat: matches 2634. .2693 of consensus"
26481. .28426
/note="L2 repeat: matches 598. .2710 of consensus"
29914. .30079
/note="MER5B repeat: matches 2. .174 of consensus"
30092. .30248
/note="MER5A repeat: matches 9. .189 of consensus"
30265. .30648
/note="L2 repeat: matches 2326. .2708 of consensus"
30717. .30795
/note="MER5B repeat: matches 64. .148 of consensus"
31403. .31535
/note="MIR repeat: matches 59. .187 of consensus"
33403. .33436
/note="L1 repeat: matches 5013. .5046 of consensus"
33437. .33728
/note="AluJo repeat: matches 1. .298 of consensus"
33729. .34269
/note="L1 repeat: matches 4472. .5013 of consensus"
34283. .34415
/note="AluJo repeat: matches 1. .135 of consensus"
34417. .34520
/note="L1 repeat: matches 4370. .4477 of consensus"
34526. .34673
/note="MIR repeat: matches 18. .157 of consensus"
35400. .35672
/note="L2 repeat: matches 1882. .2147 of consensus"
36170. .36326
/note="LTR33 repeat: matches 29. .200 of consensus"
36329. .36400
/note="L2 repeat: matches 2425. .2498 of consensus"
36524. .36755
/note="L1MC/D repeat: matches 5376. .5619 of consensus"
37157. .37425
/note="AluSg repeat: matches 28. .295 of consensus"
38277. .38503
/note="L1M4 repeat: matches 1775. .1999 of consensus"
38764. .38883
/note="60 copies 2 mer aa 61% conserved"
38897. .38945
```

Query Match 87.6%; Score 18.4; DB 9; Length 190223;
Best Local Similarity 95.0%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATGATTCACATTAGAGACT 11
 |||||
 DB 170540 CATGATTCACATTAGAGACT 170559
 |||||
 RESULT 7
 AC116458/c
 LOCUS
 DEFINITION Mus musculus chromosome 16, clone RP23-11L14, complete sequence.
 AC116458
 VERSION AC116458.9 GI:33147376
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 219241)
 TITLE Mus musculus chromosome 16, clone RP23-11L14
 JOURNAL
 REFERENCE
 2 (bases 1 to 219241)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,I., Boguslavskiy,L.,
 Boukhgalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
 Landers,T., Lechoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol.R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,J., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 219241)
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
 Macdonald,P., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol.R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,J., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
 Macdonald,P., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol.R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,J., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (23-JUL-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 5 (bases 1 to 219241)
 REFERENCE
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
 Macdonald,P., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol.R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,J., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 23, 2003 this sequence version replaced gi:32526975.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green,P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L22282
 Center clone name: 11_L14
 ----- Location/Qualifiers
 1. 219241
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="16"
 /map="16"
 /clone="RP23-11L14"
 /clone_lib="RP23-11L14" Female Mouse BAC
 1. 6
 /note="clone boundary
 clone end:SP6
 site:ECORI"
 124..227
 /rpt_family="Lx9"
 misc_feature
 source
 FEATURES
 repeat_region

```

repeat_region      244. .284
/rpt_family="(CAAAA)n"
repeat_region      308. .460
/rpt_family="Lx9"
unsure            320. .393
/note="single clone coverage"
unsure            353. .360
/note="<30 qual single clone coverage"
unsure            376. .389
/note="<30 qual single clone coverage"
unsure            complement(498. .624)
/note="single clone coverage"
unsure            complement(498. .553)
/note="<30 qual single clone coverage"
repeat_region      538. .617
/rpt_family="Lx9"
unsure            complement(565. .571)
/note="<30 qual single clone coverage"
repeat_region      735. .772
/rpt_family="AT_rich"
repeat_region      792. .854
/rpt_family="(TA)n"
repeat_region      857. .1022
/rpt_family="L1MD3"
repeat_region      1204. .1334
/rpt_family="L1MB3"
unsure            1293. .1306
/note="<30 qual SNGL region"
repeat_region      complement(1366. .2273)
/rpt_family="Lx2"
repeat_region      2346. .2397
/rpt_family="(CATATA)n"
repeat_region      2454. .3364
/rpt_family="L1"
repeat_region      complement(3366. .3929)
/rpt_family="L1MA5"
repeat_region      3938. .4218
/rpt_family="Lx4"
repeat_region      4303. .4331
/rpt_family="AT_rich"
repeat_region      complement(4332. .10730)
/rpt_family="L1P"
repeat_region      10766. .10805
/rpt_family="(T)n"
repeat_region      10842. .10865
/rpt_family="AT_rich"
repeat_region      10896. .10919
/rpt_family="AT_rich"
repeat_region      complement(11245. .11362)
/rpt_family="Lx2"
repeat_region      12557. .12780
/rpt_family="Lx9"
repeat_region      12960. .13014
/rpt_family="A-rich"
repeat_region      complement(14010. .14029)
/rpt_family="MLT1A"
repeat_region      complement(14030. .14378)
/rpt_family="ORR1B1"
repeat_region      complement(14379. .14492)
/rpt_family="MLT1A"
repeat_region      14573. .14660

```

Query Match 87.6%; Score 18.4; DB 10; Length 219241;
Best Local Similarity 95.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCATGTATCCCAATTAGAGAC 20
|||||
Db 54143 TCATGTATCCCAATTAGAGC 54124

LOCUS AC087175 22996 bp DNA linear PRI 08-OCT-2003
DEFINITION Homo sapiens BAC clone RP11-139E5 from 7, complete sequence.
ACCESSION AC087175
VERSION AC087175.6 GI:18093075
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 22996)
AUTHORS Sulston, J.E. and Wilson, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99053792
PUBMED 9847074
REFERENCE 2 (bases 1 to 22996)
AUTHORS Dignan, G., Haakenson, W., Hawkins, M. and Isak, A.
TITLE The sequence of Homo sapiens BAC clone RP11-139E5
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 22996)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 22996)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 22996)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (29-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 22996)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jan 9, 2002 this sequence version replaced gi:14530888.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
----- Center project name: H_NH0139R05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/STB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Data from AC012590 and AC073341 was used to finish this clone, AC087115.

The clone sequenced to the left is RP11-549123 the clone sequenced to the right is CTD-2375H4. Actual start of this clone is at base position 61272 of RP11-549123 actual end is at base position 98716 of CTD-2375H4.

FEATURES

Location/Qualifiers

```

source
1..22996
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-139E5"
/clone_lib="RPCI-11"
91..308
/rpt_family="MIR"
767..903
/rpt_family="L1"
1085..1423
/rpt_family="Alu"
2139..2263
/rpt_family="L2"
2694..3025
/rpt_family="Alu"
3658..3799
/rpt_family="L1"
3800..3830
/rpt_family="TTTGTG)n"
3831..4110
/rpt_family="Alu"
4111..4661
/rpt_family="L1"
4662..4956
/rpt_family="Alu"
4957..5486
/rpt_family="L1"
5637..5717
/rpt_family="MIR"
5726..6088
/rpt_family="CAT)n"
6324..6433
/rpt_family="Alu"
6434..6518
/rpt_family="CA-rich"
6519..6647
/rpt_family="L1"
7529..7803
/rpt_family="L1"
7804..8108
/rpt_family="Alu"
8109..8635
/rpt_family="L1"
8741..8871
/rpt_family="MIR"
8945..9030
/rpt_family="C-rich"
9936..9984
/rpt_family="C-rich"
10032..10904
/rpt_family="ERV1"

```

```

repeat_region 10982..11046
/rpt_family="(TG)n"
repeat_region 11403..11484
/rpt_family="MER1_type"
misc_feature 11700..12197
/note="CpG island (GC=75.3, o/e=0.83, #CpGs=58)"
repeat_region 11806..11834
/rpt_family="GC-rich"
repeat_region 13006..13238
/rpt_family="L1"
repeat_region 13265..13354
/rpt_family="Alu"
repeat_region 13357..13719
/rpt_family="MAlR"
repeat_region 13721..13959
/rpt_family="Alu"
repeat_region 14032..14292
/rpt_family="L1"
repeat_region 14305..15423
/rpt_family="L1"
repeat_region 15424..15709
/rpt_family="Alu"
repeat_region 15710..15895
/rpt_family="L1"
repeat_region 15896..16934
/rpt_family="L1"
repeat_region 17038..17171
/rpt_family="MIR"
repeat_region 17229..17687
/rpt_family="ERV1"
repeat_region 17735..17851
/rpt_family="MIR"
repeat_region 19154..19327
/rpt_family="MER2_type"
repeat_region 19328..19510
/rpt_family="Alu"
repeat_region 21845..21887
/rpt_family="A-rich"
repeat_region 22665..22961
/rpt_family="Alu"

```

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 22996;
 Best Local Similarity 90.5%; Pred. No. 1.2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 TCATGTATCCATAGACT 21
|||||
Db 16034 TCATGTATCCATAGGACT 16054

```

RESULT 9

AC068584/c 65598 bp DNA linear PRI 31-JAN-2003
 LOCUS AC068584 Homo sapiens chromosome 17, clone RP11-450A4, complete sequence.
 DEFINITION AC068584
 ACCESSION AC068584.11 GI:28173151
 VERSION AC068584.11
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 65598)
 Birren, B., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 17, clone RP11-450A4
 Unpublished
 2 (bases 1 to 65598)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
 Boguslavskiy, L., Bouckhalter, B., Brown, A., Buckett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS